An interpretive thematic analysis of the *p*-factor literature and an empirical investigation of the relationship between the *p*-factor and childhood trauma and reflective function

Laura Gibbon

D.Clin.Psy. thesis (Volume 1)

2017

University College London

Thesis declaration form

UCL Doctorate in Clinical Psychology

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:

Name: Laura Gibbon

Date: 22 June 2017

Overview

The psychiatric diagnostic classification system is the dominant conceptual framework within which clinical practice and mental health research are conducted. Recent research has identified the *'p*-factor', a tendency towards experiencing symptoms of psychopathology comorbidly (Caspi et al., 2014), which is part of a broad attempt to develop an empirical nosology for psychopathology. Part one of this thesis is a systematic review which critically analyses this body of quantitative research, using a reviewing method adapted from thematic analysis (Braun & Clarke, 2006).

Empirical nosological research into Axis I and personality disorder symptomatology has tended to be conducted separately. However, the '*p*-factor', describing Axis I comorbidity, and a 'general personality disorder factor' (Sharp et al., 2015) have both been extracted, raising the questions of how they relate to one another and whether they reflect psychopathological severity. Part two of this thesis is an empirical paper comparing alternative models of the comorbidity structure of internalising, antisocial, thought disorder and borderline symptoms, and the relationship between the extracted *p*-factor and childhood trauma and reflective function.

Research attempting to establish an empirically-grounded nosology for mental health employs complex statistical techniques and requires access to large amounts of comprehensive data, which may make it difficult for clinicians to undertake. However, this influential area of research has potentially significant implications for mental health practice. Part three of this thesis is a critical appraisal of the research process, with particular reference to the ways in which a clinical perspective might be important for this type of research.

Table of contents

List of tables and figures

Literature review

Tables

Figures

Empirical paper

Tables

Figures

Acknowledgements

Firstly, I would like to thank my supervisors, Professor Peter Fonagy and Dr Tobias Nolte, for their insightful guidance, encouragement and generosity.

The Probing Social Exchanges Study would not exist without the participants who gave up their time to take part, and I would like to sincerely thank them all. The study is a collaborative effort between a large number of people: Professor Read Montague and his group at Virginia Tech Carillon; and Peter, Tobias, Dr Janet Feigenbaum and the rest of the team in London. The study research assistants, including Natasha Smyth, Hinal Patel, Isla Fichie, Clarissa Bauer-Staeb and Julia Griem, have collecting much the data I used. I am particularly grateful to Natasha, Clarissa and Julia for coordinating my participation in the study. I would also like to thank Ravi Das for discussing with me ways of dealing with missing data in Mplus.

Finally I would like to thank my mother and sister, Charlotte and Anne, my friends and, particularly, Kathleen, Magnus and Alec.

Part 1: Literature Review

What is the *p*-factor, how radical a challenge does it pose to the diagnostic classification system and what are its clinical implications?

A thematic analysis

Abstract

Aims: The '*p*-factor', a statistical construct reflecting a tendency towards experiencing symptoms comorbidly, has been extracted as part of a broad empirical nosology programme. This has potentially significant ramifications for the diagnostic classification system. This review aims to characterise the *p*-factor, to determine how far it challenges the diagnostic paradigm and to consider its clinical implications.

Method: A systematic search of three databases was conducted and studies were screened according to the review inclusion criteria. This review piloted a method adapted from thematic analysis (Braun & Clarke, 2006) for the purposes of reviewing quantitative studies.

Results: Fourteen papers met the inclusion criteria for the review, including studies using child, adolescent, community adult and psychiatric adult samples. An integrative synthesis of the quantitative data identified differences between these studies. An interpretive thematic analysis based on pre-determined themes identified several respects in which the *p*-factor research retains assumptions of the diagnostic system.

Conclusions: The results of the integrative and interpretive syntheses were brought together in an integrated discussion around the review questions, identifying how inherited assumptions of the diagnostic system might influence conceptions of its clinical implications.

Introduction

The diagnostic system and its critics

The psychiatric diagnostic classification system, first proposed by Emil Kraepelin at the end of the nineteenth century, is the dominant conceptual framework within which clinical practice and mental health research are conducted. Kraepelin extrapolated from medicine to suggest that particular 'disease processes' would be likely to share symptoms, a biological basis and aetiology (Bentall, 2004). Various criticisms of the classification system have been mounted and the weight of such arguments is widely acknowledged (Beutler & Malik, 2002), however, although the diagnostic system has gone through a series of iterative developments (American Psychiatric Association, 1980; 1987; 1994; 2000; 2013), the notion of discrete diagnostic categories defined by operationalisable criteria has persisted.

The conceptual unity afforded by the diagnostic system is seen by some as an important virtue and proponents point to its utility as a pragmatic indicator for treatment, as well as its role in service organisation and clinical communication (Jablensky, 2016). Even among those who note its problems, it has been suggested that the challenge is to ensure our diagnoses keep up with scientific developments, insofar as this is possible. Others have claimed that research findings pose a more significant challenge, and that the diagnostic system is not a source of desirable unity but rather of stagnation in our understanding and treatments (Cuthbert & Insel, 2013; Kinderman, Read, Moncrieff, & Bentall, 2013). A widely-shared hope is that grouping individuals in more clinically-meaningful ways could improve matters, and to this end the National Institute for Mental Health has ceased funding to research projects which use the diagnostic framework and founded the Research Domain Criteria programme (RDoC), which aims to identify dimensions of behaviour which are subsumed by identified neural circuits (T. Insel et al., 2010). The diagnostic system has also been challenged on conceptual grounds, on the basis that phenomenologically-diagnosed disorders are defined by their symptoms but are also taken to be the cause of those

symptoms (Kinderman et al., 2013); for example, ‘anxiety’ is both a symptom and the disorder it signifies.

Scientific paradigms

Thomas Kuhn’s account of scientific knowledge production was based on his study of the history of science, which he argued was characterised by lengthy periods of ‘normal science’, conducted within established paradigms, interspersed with episodic paradigm shifts (Kuhn, 1996). Paradigms are a set of background assumptions about the world which determine the focus of scientists, and scientists working within a paradigm generally agree on answers to questions including what entities exist, how they interact, how they can be measured, what questions can be asked about those entities, what techniques can be used to answer these questions and what counts as evidence (Ladyman, 2001). An important component of a paradigm is the ‘disciplinary matrix’, a set of explicit and implicit answers to these questions, as well as practical skills, tacit knowledge and general attitudes shared by scientists. Science conducted within an established paradigm is focussed on elaborating on previous findings, is ‘puzzle-solving’, and is conservative insofar as fundamental paradigmatic assumptions are not questioned. When empirical or conceptual anomalies accumulate and assumptions are challenged, a paradigm shift may follow, during which the conservative puzzle-solving mentality is replaced with a problem-solving approach, with conceptual problems requiring more creative solutions (Ladyman, 2001).

The diagnostic system paradigm

Kuhn’s ideas have been enthusiastically taken up in the social sciences, despite the fact he developed the concept of a paradigm by noting features of the natural sciences which were missing in social science; it was the lack of features such as an almost universally-shared set of background assumptions and an exemplar experiment which led him to the conclusion they were pre-paradigmatic (Fuller, 2000). However, in the field of mental health the diagnostic system has conferred a degree of ontological and pragmatic unity beyond what is typical of a study of human experience and behaviour. Multiple factors

are known to be related to psychopathology and research is pluralistic, but the diagnostic framework provides a unifying translational framework. The recent NIHR funding changes and RDoC notwithstanding, research funding, clinical trials, population studies, pharmacological and psychological interventions, organisation of services and clinical guidelines tend to be structured around diagnoses.

Regarding the ontological questions of what entities exist and how these entities interact, in many respects the classification system remains true to Kraepelin (Craddock & Owen, 2010). Mental health disorders are conceptualised similarly to physical diseases as dichotomous, representing a state of ill-health, and sharing an aetiology, biological basis and symptom profile (Bentall, 2004). Disorders are conceptualised as unobservable latent entities, which are independent of one another and which give rise to symptoms (Kendler, 2016). The view that we should believe in unobservable theoretical entities is scientific realism (Ladyman, 2001), a complex position which proponents of the diagnostic system may not necessarily be committed to, however, pragmatic realism about disorders appears to be an important assumption of the paradigm.

Epistemological questions include what research questions can be asked about existent entities and what counts as evidence within the paradigm, as well as the methodological questions of how entities can be measured and what techniques can be used to answer research questions about them (Ladyman, 2001). In psychology there is an epistemological gap between unobservable objects of interest, such as disorders, and what can be measured, such as behaviour or symptoms (Essex & Smythe, 1999), and an important task of research is determining the reliability and validity of measures of psychological constructs. As symptoms are understood to follow from disorders, improved symptom characterisation is considered to lead to a superior nosology (American Psychiatric Association, 2013). Quantifying between-person differences, using statistical methods, is the paradigmatic way of gaining knowledge of disorders.

The general psychopathology factor

Moving beyond early research investigating how disorders cluster together (Caron & Rutter, 1991; Sturt, 1981), the availability of population data and statistical modelling techniques have made data-driven investigations of psychopathology possible. Factor analytic approaches aim at better measurement of variables which, either for practical or conceptual reasons, cannot be measured directly. A robust finding across different groups is the existence of 'internalising' and 'externalising' dimensional factors, which represent the degree to which symptoms within a dimension tend to be associated (Krueger & Markon, 2011). This offers a way of interpreting phenomena including that there are shared genetic and environmental risk factors for different mental health presentations and that specific biomarkers for disorders have been difficult to identify. Lahey and colleagues (2012) observed that dimensional latent factors were themselves correlated, and mooted the possibility of a general propensity to psychopathology. The robustness of the internalising and externalising structure precipitated a move from exploratory to confirmatory modelling techniques, which now predominate, and which were extended by Lahey and colleagues to include bifactor models, which comprise a higher-order general factor in addition to 'group' factors.

Caspi and colleagues (2014) were then the first to thoroughly investigate this hypothesis in a longitudinal dataset which included measures of a broad range of symptoms including thought disorder.[1] They found that Axis I symptomatology was best described by a bifactor model, comprising internalising and externalising group factors and a higher-order general psychopathology factor on which thought disorders loaded, coining the term 'the *p*-factor' to describe the "tendency to experience psychiatric problems as persistent and comorbid" (Caspi et al., 2014; p. 131). That thought disorders did not form a viable group factor influenced the authors' 'structural hypothesis' that internalising and externalising

[1] There is parallel research into the structure of personality disorders and a general factor had previously been extracted using exploratory bifactor modelling (Wolf, Miller, & Brown, 2011).

disorders represent pathological expressions of gendered personality styles, with characteristics of high-*p* individuals leading them to experience persistent difficulties, likely to be associated with thought disorder. The *p*-factor invites a different way of thinking about psychopathology and it may have important implications for research, treatment, service organisation and efforts at prevention.

The current review

It is not clear whether the *p*-factor warrants a tweaking of the conceptual framework or whether it implies deeper conceptual problems with the diagnostic system. Firstly, the *p*-factor is a statistical construct and it is not clear how it should be interpreted, either at the population- or individual-level, or what its clinical implications might be. Secondly, it is not clear to what extent models of the comorbidity structure of psychopathology challenge broader assumptions of the diagnostic paradigm, or what the statistical assumptions of sophisticated modelling techniques might imply. This review will aim to answer these questions by examining Caspi and colleagues' (2014) seminal paper (hereafter referred to by its study identification 'CASPI2014') and research it has stimulated which aims to extract statistical constructs analogous to the *p*-factor. For clarity, any general factor extracted within a bifactor model structure will be referred to as a '*p*-factor', whether or not the study authors used this terminology.

Review questions

1. What is the *p*-factor?
2. Does the *p*-factor challenge the diagnostic classification system paradigm and, if so, in what ways?
3. What are the clinical implications of the *p*-factor?

Method

Identification and selection of studies

Search strategy

The aim of this review differs from more typical literature reviews and the search strategy reflected this fact. Rather than seeking to identify a range of papers investigating a phenomenon, this review sought to identify studies adopting a particular methodology (confirmatory factor analysis; CFA), which were motivated by CASPI2014 to investigate the *p*-factor. As the focus of the review was specific in this regard, the search strategy was restricted to text words rather than subject areas. Databases PsychInfo, Web of Science and SCOPUS were searched for English language papers published using the search terms ‘p factor’ and ‘general psychopathology’, published between 2013 (when CAPSI2014 became available online) and September 2016.

Supplementary hand searching was conducted using Google Scholar metrics to identify all papers which cited CASPI2014 up to and including September 2016. Following the identification of studies meeting the inclusion criteria, the reference lists of these papers were searched for additional relevant studies. The papers identified at each stage are shown in Figure 1.

Inclusion and exclusion criteria

Studies were included in the review according to the following five criteria.

Table 1

Review inclusion and exclusion criteria

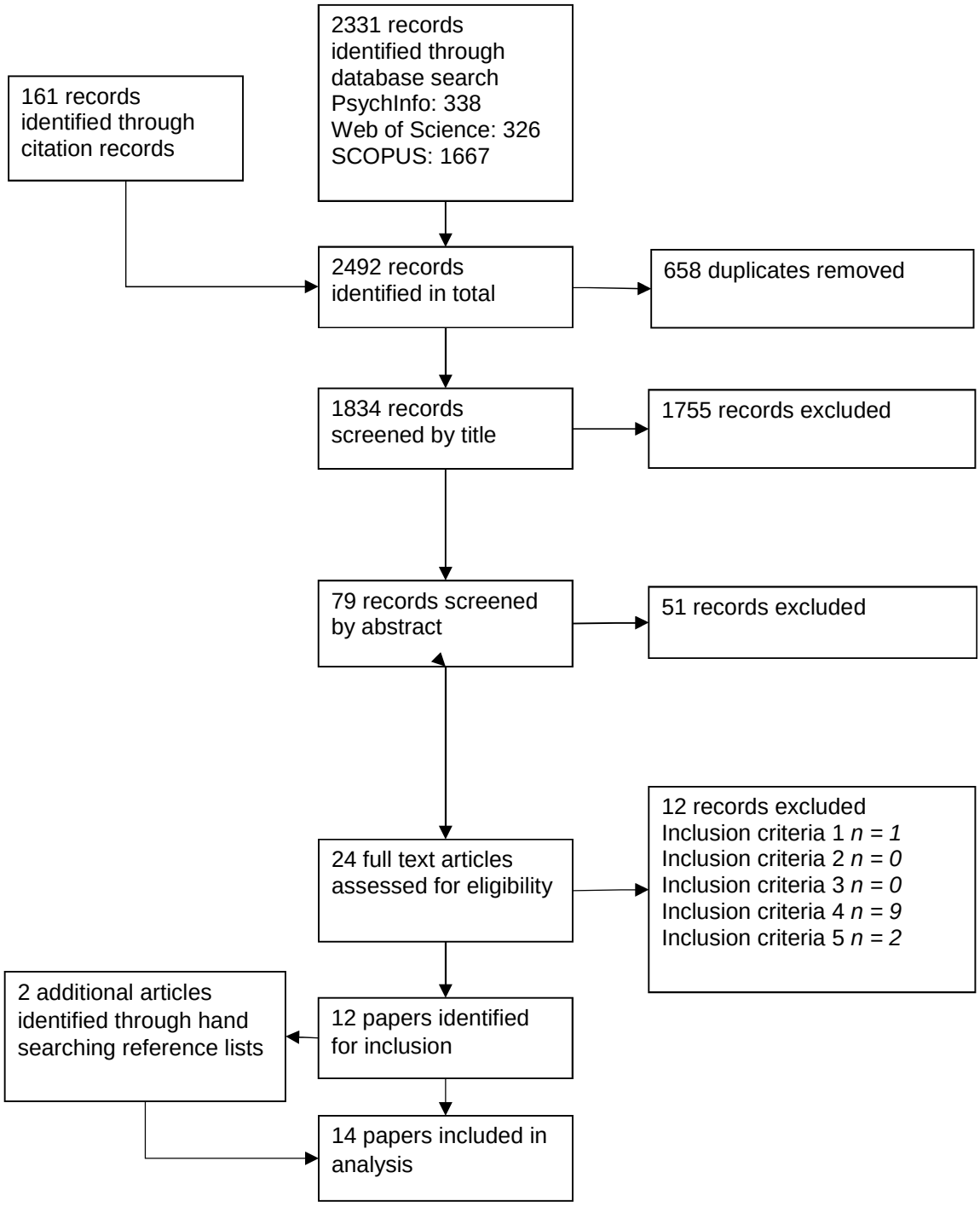
Inclusion criteria	Included	Excluded
1. Type of study	Quantitative primary research studies using real data	All other research studies and papers. For example: studies using simulation data; qualitative studies; secondary research studies; commentary papers
2. Publication type	Published in a peer-reviewed journal	All other types of publication. For example: books, dissertations
3. Language	English language	All other languages
4. Aims, design and method	Used CFA to either extract the *p*-factor or investigate the comorbidity structure of psychopathology with at least one bifactor model. Compared at least two models of psychopathology using CFA. Motivated by CASPI2014 (discussed in the introduction as motivating the study hypotheses).	*No exclusions for*: the use of supplementary exploratory or structural equation modelling methods, provided the CFA criterion is met
5. Population and measures	Appropriately broad symptom coverage, such that: *either* the range of symptoms measured *or* the sample population included symptoms relating to at least two of internalising, externalising or thought disorder (the group factors identified by CASPI2014)	*No exclusions for*: age of participants; additional types of symptomatology measured (for example, Axis II or neurodevelopmental)

Identification of relevant studies

In order to determine whether papers met the inclusion criteria they were first screened by title and then, where necessary, by abstract, using Endnote X7.7.1 (Thompson Reuters, 2016). Full text copies of the remaining papers were obtained and checked against the inclusion criteria. This process is summarised in the flowchart shown in Figure 1.

Figure 1

Flow chart showing included studies



Method of analysis

An interpretive review of quantitative evidence: rationale for a novel review strategy

Noblit & Hare (1988) make a distinction between *integrative* literature reviews of empirical research (which, on a Kuhnian understanding, reflect the assumptions of the relevant paradigm) and *interpretive* literature reviews of qualitative research. The review questions posed here were interpretative, requiring critical consideration of the assumptions inherent in the studies. Although all reviews of empirical evidence will inevitably involve aspects of interpretation (Noblit & Hare, 1988), for this to be an explicit part of the method a novel strategy was required. Theoreticians of science, including Kuhn, have drawn attention to the value-laden nature of the scientific process (Ladyman, 2001), which in psychology involves formulating a research question in natural language, collecting and analysing data according to paradigmatic methods (which have their own theoretical assumptions) and then interpreting findings, again, in natural language (Mareschal, 2007). Research papers therefore include quantitative and qualitative (text) data, and therefore an objective of this review was the synthesis of both types of data so that they could be analysed in relation to the review questions.

Adaptions to thematic analysis

The objectives of the review necessitated a flexible qualitative approach which could be adapted in order to incorporate quantitative data, and Braun and Clark's (2006) method of thematic analysis was selected due to its epistemological flexibility. The first stage in the synthesis was extraction and tabulation of the quantitative data relevant to the review questions. Questions relating to study quality were outside the scope of this review and there are no quality assurance tools available for CFA factor structure studies, so this was not considered.

In order to answer the question of how far this body of research challenges the diagnostic paradigm, it was necessary to formalise the assumptions of the diagnostic paradigm which had been identified in the introductory review of relevant literature, and

these are shown in Table 2. It was noted that during periods of normal science, key paradigmatic assumptions are taken to be evidently true (Ladyman, 2001) and so may not be explicitly stated in the reviewed papers.

The syntheses of data through the generation of comparison tables of quantitative data and the qualitative thematic analysis then informed an integrative discussion around the three review questions. Importantly, all of the review questions depended on an analysis of both the quantitative and qualitative data. A summary of how the objectives of this review were met through adaptions to the six-phase Braun and Clarke (2006) method are detailed in Table 3.

Table 2

Pre-determined themes for thematic analysis

Ontological assumptions	Consequent epistemological assumptions
Disorders as discrete categories	*Scope of research*
Mental health disorders are categorical, dichotomous entities	There is no consistent approach to defining severity of presentation
Mental health disorders are independent; a person diagnosed with one disorder will not be more likely to be diagnosed with another disorder	Improved knowledge of disorders follows from observation of 'pure' disorders; that is, comorbid presentations obscure understanding
Realism about diagnoses	*Appropriateness of statistical methods*
Mental health disorders exist; that is, there are entities (e.g. 'depression') which exist across people	Knowledge about mental health disorders can be gained by examining between-person factors
Diagnoses as latent entities	*Observation and measurement*
Mental health disorders are unobservable latent entities	Mental health disorders cannot be known about directly, but only through examination of observable signs
Symptoms are observable signs of latent disorders	Symptoms of a particular disorder should reliably and validly specify that disorder
	Improved symptom characterisations will improve knowledge of mental health disorders
Symptoms follow from a disorder and are independent of one another	Unidirectional causation should be expected (cause -> disorder -> symptom); interactions between symptoms should not be expected
Key characteristics of disorders	*Focus of research*
A mental health disorder will have a specific aetiology, biological basis and symptom profile	It should be possible to identify biomarkers and aetiological risk factors for specific disorders
	Aetiology and the biological basis of disorders are informative for natural course
A mental health disorder will respond to a specific treatment(s)	Intervention research should group people by disorder
	Aetiology and the biological basis of disorders are informative for treatment choice and outcomes

Table 3

Adaption of Braun and Clarke's (2006) thematic analysis method

Braun and Clarke thematic analysis method	Adaptions for this review
Phase 1: Familiarising yourself with your data	***Phase 1: Familiarisation with quantitative and qualitative data***
Reading the data	Qualitative data: reading the papers; including data relevant to the themes, and how the *p*-factor and its clinical implications are interpreted. Quantitative data: examination of hypotheses, design, method and findings for comparison. The key points of comparison were tabulated and briefly summarised to support subsequent phases of the analysis.
Phase 2: Generating initial codes	***Phase 2: Searching for pre-identified themes and coding data***
Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code	The Braun and Clark (2006) method is flexible regarding whether themes are pre-determined or data-driven; here they were pre-determined according to the review questions. As themes were pre-determined, the data were coded accordingly and this phase involved synthesising of data around these themes. Themes were based on the identified characteristics of the diagnostic paradigm (Table 2). According to the Braun and Clark method, themes can be semantic or latent. As the themes related to explicit and implicit paradigmatic assumptions, the themes were considered to generally be both semantic and latent.
Phase 3: Searching for themes	
Collating codes into potential themes, gathering all data relevant to each potential theme	
Phase 4: Reviewing themes	
Checking in the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic 'map' of the analysis	
Phase 5: Defining and naming themes	
Ongoing analysis to refine the specifics of each theme, and the overall story the analysis tells; generating clear definitions and names for each theme	
Phase 6: Producing the report	***Phase 3: Reporting the thematic analysis***
The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis	The emphasis in this review was on a narrative synthesis around the themes. Due to the large amount of qualitative and quantitative data, it was important that quotes were used sparingly, with a focus instead on citing studies where they were relevant to one of the questions raised.
	Phase 4: Integrative discussion
	The review questions were answered in an integrative discussion, drawing on both the initial familiarisation with the qualitative and quantitative data, and the results of the thematic analysis

Results

Details of included studies

Fourteen studies met the inclusion criteria. One study replicated CASPI2014 in a different population [LACEULLE2015 (Laceulle, Vollebergh, & Ormel, 2015), four studies aimed to extract the *p*-factor [CARRAGHER2016 (Carragher et al., 2016); LAHEY2015 (Lahey et al., 2015); MARTEL2017 (Martel et al., 2017) and PATALAY2015 (Patalay et al., 2015)], and eight aimed to model the factor structure of psychopathology more broadly but tested at least one bifactor model [BRODBECK2014 (Brodbeck et al., 2014); CASTELLANOS2016 (Castellanos-Ryan et al., 2016); HOERTEL2015 (Hoertel et al., 2015); NOORDHOF2015 (Noordhof, Krueger, Ormel, Oldehinkel, & Hartman, 2015); STOCHL2015 (Stochl et al., 2015); SUBICA2015 (Subica, Allen, Frueh, Elhai, & Fowler, 2015) and WALDMAN2016 (Waldman, Poore, van Hulle, Rathouz, & Lahey, 2016)]. Finally, KIM2015 (Kim & Eaton, 2015) aimed to determine whether exploratory factor models derived using the 'Bass-Ackwards' method (Goldberg, 2006) were comparable with a bifactor model of the *p*-factor.

Phase 1: Familiarisation with the quantitative data

Hypotheses and design

CFA is a type of structural equation modelling (SEM), in which factors are assumed to be independent. CFA models are considered to be hypotheses to explain the data and, typically, alternative models are specified and then a model is selected on the basis of fit statistics and qualitative factors including model parsimony, interpretability of factors and clinical relevance (Geiser, 2013). A CFA 'measurement model' may then be extended to become a full SEM, in which causal paths between factors are hypothesised. In all study characteristic tables, data relates to the CFA measurement model only and not any additional study hypotheses for which different data were used.

The studies which specifically aimed to extract the *p*-factor tended to test models robustly identified by previous studies, albeit in some cases with novel group factor structures (NOORDHOF2015; Table 5). Several of the studies used exploratory methods either prior to CFA (BRODBECK2014), in relation to broader study aims (CARRAGHER2015; KIM2015) or to make *post-hoc* modifications to a model (CASPI2014).

Sample population

The *p*-factor is a statistical construct which relates to variability in a population; an instructive analogy is the construct of 'heredity' in quantitative genetics (Plomin, DeFries, Knopik, & Neiderheiser, 2013), and all of the study authors noted their findings related only to their sample population. Sample characteristics are given in Table 4; in all tables and figures, data from child/adolescent and adult populations are presented separately.

Nine studies used samples from a child or adolescent population, all of which were community samples. Five studies used samples from an adult population, two of which were psychiatric and three of which were community samples, of which two were from the same population cohort study (HOERTEL2015 and KIM2015; Table 4). The studies used samples from 10 different countries, however, only one (MARTEL2017) was from a country outside Europe or the United States. The community samples varied in the breadth of their coverage. The psychiatric samples were drawn from different clinical populations, an important difference as comorbidity is more common in people with severe and enduring mental health problems (Kessler, Chiu, Demler, & Walters, 2005).

Table 4

Details of sample used to estimate CFA measurement models

Study ID	Population	Cohort study name	Country	Sample size	Mean age (sd) in years
Child/ adolescent population					
CARRAGHER 2016	Community adolescent	CAP	Australia	N = 2175	13.3 (0.48)
CASTELLANOS 2016	Community adolescent	IMAGEN	Ireland, Germany, France, UK	2 time points: T1: N = 2144 T2: N = 1603	T1: 14.39 (0.77) T2: 16 (sd not reported)
LACEULLE 2015	Community adolescent	TRAILS	The Netherlands	4 time points T1: N = 2230; T2, 3, 4: not stated	T1: 10.5 (0.58) T2: 13.6 (0.59) T3: 16.1 (0.59) T4: 19.1 (0.60)
LAHEY2015	Community child/ adolescent	Pittsburgh Girls Study	US	N = 2450	Not reported Range: 5-11 years
MARTEL2017	Community child	High Risk Cohort Study for Psychiatric Disorders	Brazil	Parent interview N = 2512 Child evaluations N = 2395	9.65 (1.93)
NOORDHOF 2015	Community adolescent	TRAILS	The Netherlands	3 time points T1: N = 2230 T2: N = 2150 T3: N = 1815	11.09 (0.55)
PATALAY2015	Community adolescent	Me and My School	UK	N = 23,477	12.05 (0.56)
STOCHL2015	2 samples: 1. Community adolescent 2. Community adolescent	1. ALSPAC 2. ROOTS	1. UK 2. UK	Sample 1: N = 6617 Sample 2: N = 977	Sample 1: 13 Sample 2: 18 (sd not reported)
WALDMAN 2016	Community child/ adolescent	Tennessee Twin Study	US	N = 3136 (1568 twin pairs)	11.7 (3.3)

Study ID	Population	Cohort study name	Country	Sample size	Mean age (sd) in years
Adult population					
BRODBECK 2014	Psychiatric adult (outpatient; excluding psychosis, mania, substance misuse)	n/a	Switzerland	N = 1024	39.69 (14.62)
CASPI2014	Community adult	Dunedin Study	New Zealand	5 time points T1: N = 1037 T5: N = 957 (other time points not reported)	T1: 18 T2: 21 T3: 26 T4: 32 T5: 38 (sd not reported)
HOERTEL2015	Community adult	NESARC	US	2 time points N = 34,653 (T1 and T2; people with missing data at T2 excluded)	Mean not reported. Range: T1: 18 to <90 T2: 20 to <90
KIM2015	Community adult	NESARC	US	2 time points: T1: N = 43,093 T2: N = 34,653	Mean not reported. Range: T1: 18 to <90 T2: 20 to <90
SUBICA2015	Psychiatric adult (inpatient severe; 31.6% diagnosed with a personality disorder)	n/a	US	N = 962	36.66 (14.82)

Abbreviations: ALSPAC: Avon Longitudinal Study of Parents and Children; CAP: Climate Schools and Preventure; NESARC: National Epidemiologic Survey on Alcohol and Related Conditions; TRAILS: TRacking Adolescents' Individual Lives Survey

Data

Type of data

Two of the studies used data on diagnosis (without exclusion rules), two used likelihood of diagnosis, seven used symptom data defined by measure subscales and three used item-level symptom data. Data were then treated as categorical (dichotomous or ordinal) or continuous (Table 5). As noted by CARRAGHER2016, symptom-level modelling can help to identify low prevalence conditions.

Cross-sectional and repeated measures data

Two studies used repeated measures data to test their CFA measurement models (Table 5), thereby reflecting the difference between persistent and episodic presentations, which as CASPI2014 notes differ in their aetiology and course. Several other studies used longitudinal data to test additional hypotheses, although these are not reviewed here.

Range of symptomatology

There were symptoms measured across various group factors including internalising (INT) and its sub-factors fear (FEAR) and distress (DIST), externalising (EXT), thought disorder (TD), autism spectrum (ASD), attentional and orientation difficulties (ATT-OR) and additional disorder-type group factors. With the exception of HOERTEL2015, personality disorder symptomatology was not measured. There was considerable range in symptomatology measured, both within and between group factors. The majority of studies considered symptoms to be caused by the same group factors, however there were exceptions (for example, LAHEY2015's treatment of depression; Table 5 and Figure 2). There were also symptoms measured which have been less commonly modelled in the literature, and these were treated differently by different studies (for example, attentional difficulties by NOORDHOF2015 as opposed to WALDMAN2016 and LAHEY2015; Table 5). Finally, there were differences in how much data was used to specify the factors (Table 5); as CARRAGHER2016 notes, specification by a restricted symptom range decreases the likelihood of identifying a robust factor.

There were several specific issues relating to symptomatology measured. One study (SUBICA2015) met the inclusion criteria due to the sample population having severe psychiatric presentations, as only a narrow range of symptoms (depression and anxiety) were measured. Therefore in this case the general latent factor modelled is more akin to INT than *p*, although there was considerable unmeasured comorbidity in the sample (Table 4). A broader, but nonetheless restricted, set of symptoms was measured by STOCHL2015. Finally, despite measuring the symptoms of paranoia and psychosis (Table 5), BRODBECK2014 used a psychiatric sample in which these presentations were excluded (Table 4).

Method of collection

Nine studies used self- or parent-report measures to collect data and four used clinical interviews (Table 5). Although it was outside of the scope of this review to consider the measures used, it was noted that the appropriateness of the measures for the purposes of assessing clinical presentations varied; for example, between interviews conducted by health professionals (CASPI2014) and measures designed for screening purposes (STOCHL2015; SUBICA2015).

Table 5

Symptomatology data used to estimate CFA measurement models

Study ID	Level of data modelled	Treated as	Data collected	Grouping of symptoms[+]	Method of data collection
Child/ adolescent population					
CARRAGHER 2016	Symptom (item level)	Dichotomous	Cross-sectional	Item-level data, by factor: EXT; INT; TD	Self-report scales
CASTELLANOS 2016	Likelihood of diagnosis	Continuous	Cross-sectional (two models at each age)	EXT: ADHD; CD; ALC; DRUG; TOB INT: ANX; DEP; SAD; P/PAN; ED; OCD	Structured interview
LACEULLE2015	Symptom dimensions (subscale score)	Continuous	Repeated measures	EXT: agg; att; asoc INT: a/dep; w/dep; gad; sad; sep; pan *p: td; ocd; psy	Self-report and parent-report scales
LAHEY2015	Symptom dimensions	Continuous	Cross-sectional	EXT: cd; opp; imp; att INT: dep; gad; sad; sch; pan/som; sep	Self-report and parent-report scales
MARTEL2017	Likelihood of diagnosis	Continuous	Cross-sectional	EXT: CD; ODD; ADHD; ASD FEAR: PAN; AGOR, SAD; SEP; ANX DIST: DEP; GAD; OC; TIC; PTSD; ED	Structured interview
NOORDHOF 2015	Symptom dimensions (sub-scales)	Continuous	Cross-sectional	EXT: agg; rule INT: dep; anx; som AO: rule; att; as-u; as-o AS: as-b; as-c; as-o; as-s; as-r; as-u	Parent-report scales
PATALAY2015	Symptom (item data)	Categorical	Cross-sectional	Item-level data, by factor: EXT; INT	Self-report scales
STOCHL2015	Symptom	Continuous	Cross-sectional	INT: hall; del; td TD: dep; anx	Self-report scales; structured interview
WALDMAN 2016	Symptom dimensions	Continuous	Cross-sectional	EXT: opp; cd; imp; att INT: sad; phob; agor; sep; ocd *p: dep; gad	Self-report and parent-report scales

Study ID	Level of data modelled	Treated as	Data collected	Grouping of symptoms+	Method of data collection
Adult population					
BRODBECK 2014	Symptom (item-level)	Categorical	Cross-sectional	Item-level data, by factor: DEP; PHOB; AGG; SUI; NERV; SOM; INFO; IS	Self-report scale
CASPI2014	Disorder/ symptom counts	Categorical	Repeated measures	EXT: ALC; CAN; DRUG; TOB; CD INT: MDD; GAD; FEAR *p: OCD; BD; SCHIZ	Structured interview
HOERTEL2015	Diagnosis	Categorical	Cross-sectional	EXT: ALC; DRUG; TOB; GAMB; ASPD INT I: MDD; DYST; GAD INT II: PAN; SAD; PHOB; BD; APD; DPD; OCPD; PPD; SCPD; HPD	Structured interview
KIM2015	Diagnosis	Categorical	Cross-sectional	EXT: ALC; CAN; DRUG; TOB; ASPD DISTRESS: MDD; DYST; GAD FEAR: PAN; SAD; PHOB *p: BD	Structured interview
SUBICA2015	Symptom (item data)	Categorical	Cross-sectional	Item-level data, by factor: DEP; ANX	Self-report scales

Key: * indicates symptom or disorder only loaded on *p* in the best-fitting model; + indicates group structure of the best-fitting model where this varied across models tested

Abbreviations: Latent factors (upper-case lettering and underlined): AGG: aggression; ANX: anxiety; ATT-OR: attention-orientation; problems; ASD: autism spectrum; DEP: depression; DIST: distress; EXT: externalising; FEAR: fear; INFO: information processing; INT: internalising; IS: interpersonal sensitivity; NERV: nervous tension; PHOB: phobic fear; SOM: somatic problems; SUI: suicidal ideation. Diagnoses (upper-case lettering): ADHD: attention deficit hyperactivity disorder; ALC: alcohol dependence; ANX: anxiety; ASPD: Antisocial Personality Disorder; APD: Avoidant Personality Disorder; BD: bipolar disorder or mania; CAN: cannabis addiction; CD: conduct disorder; DEP: depression; DPD: Dependent Personality Disorder; DRUG: drug addiction (hard drugs); DYST: dysthymia; ED: eating disorder; FEAR: fear; GAD: generalized anxiety disorder; GAMB: gambling addiction; HPD: Histrionic Personality Disorder; MDD: major depression; OCD: obsessive compulsive disorder; OCPD: obsessive compulsive personality disorder; PAN: panic disorder; PHOB: specific phobia; P/PAN: phobia/panic disorder; PPD: Paranoid Personality Disorder; SAD: social anxiety disorder; SEP: separation anxiety; SCHIZ: schizophrenia; TOB: tobacco addiction. Symptoms (lower-case lettering): a/dep: anxious depression; agor: agoraphobia; as-b: behaviour and emotions not tuned to social situation (autistic spectrum); as-c: stereotyped behaviour (autistic spectrum); as-o: orientation-problems in time, place, or activity (autistic spectrum); as-s: reduced contact and social interests (autistic spectrum); as-r: resistance to change (autistic spectrum); as-u: difficulties in understanding social information (autistic spectrum); asoc: antisocial behaviour/ delinquency; att: attentional difficulties' cd: conduct disorder; del: delusions; fear: combined symptoms

of phobia, SAD and agoraphobia; hall: hallucinations; imp: hyperactivity/impulsivity; is: interpersonal sensitivity; opp: oppositional defiant; pan/som: panic and somatic symptoms; psy: psychotic experiences; rule: rule-breaking; som: somatic symptoms; sad: social anxiety; sch: school phobia; sep: separation anxiety; td: thought disorder; w/dep: withdrawn depression

Models tested

As noted above, the aims of the studies varied and this was reflected in the models they tested. Studies which based their CFA on previous exploratory work or had additional study aims tended to test a restricted set of models (e.g., BRODBECK2014; HOERTEL2015; KIM2015), whereas those basing their CFA models on previous literature tended to test similar models. HOERTEL2015, the only study to incorporate personality disorder symptomatology, based their group factor structure (Table 5) on a previous exploratory analysis of the same community sample (Blanco et al., 2013). NOORDHOF2015 allowed cross-loadings in their bifactor model (Table 5). It was not possible to describe all models tested, but the main comparison models are shown in Table 6 (with group factor structure shown in Table 5).

The studies differed according to whether they specified group factor and bifactor models with correlated (oblique) and/or uncorrelated (orthogonal) group factors. (In bifactor models, group factors can be correlated with each other but not the general factor.) Orthogonal bifactor models tend to increase the strength of *p* (Murray et al, 2015) and also preclude describing the relationships between group factors. WALDMAN2016 explored the most comparisons, that is, oblique and orthogonal variations of several alternative group and bifactor structures. Finally, only two studies tested second-order models, in which a general factor is considered to influence the group factors rather than directly influencing the manifest variables themselves. Both these studies found support for the bifactor over the second-order models, however, it should be noted that fit statistics tend to favour bifactor models (Mansolf & Reise, 2017).

Table 6

Models specified

Study ID	Models tested	Uni-factorial	Group Factors		Bifactor			Second-order	Other
			Oblique	Orthogonal	Oblique	Orthogonal	Modified		
Child/adolescent									
CARRAGHER2016	5	x			x	x			x
CASTELLANOS2016	6	x	x		x	x		x	x
LACEULLE2015	4	x	x		x		x		
LAHEY2015	3		x		x			x	
MARTEL2017	4	x	x			x			x
NOORDHOF2015	3+					x			x
PATALAY2015	3	x	x			x			
STOCHL2015	4	x	x	x		x			
WALDMAN2016	9								x
Adult									
BRODBECK2014	2								
CASPI2014	4	x	x		x		x		
HOERTEL2015	2					x		x	
KIM2015	2*					x			x*
SUBICA2015	3	x	x			x			

Key: + NOORDHOF2015 compared their bifactor model with higher-order models derived from EFA; * KIM2015 compared their bifactor model with a Bass-Ackwards hierarchy

Study results

Best fitting models

Fit statistics have limited meaning across different studies (Geiser, 2013) and so are not reviewed here. The conventions for fit of bifactor model are less established (Reise, 2012), however, according to standard conventions all of the best-fitting models were adequate (Appendix A2). A number of the studies found equivocal (e.g., BRODBECK2014; CARRAGHER2016; CASPI2014; HOERTEL2015; KIM2015; LACEULLE2015) or marginal (e.g., PATALAY2015) fit between at least two models. BRODBECK2014 concluded that both competing models fit their data adequately well, CARRAGHER2016's competing models were both bifactor, and the remaining studies erred towards endorsing their bifactor model based on factors such as parsimony.

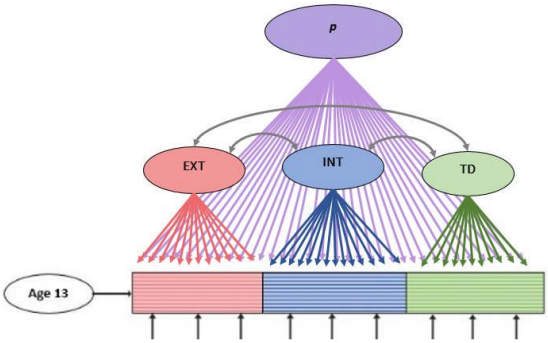
The best-fitting model or the model endorsed by the study authors (or, in the case of BRODBECK2014 and HOERTEL2015, the contending bifactor model), are shown in Figure 2. Where disorder-level latent factors were modelled these are shown (CASPI2014 and LACEULLE2015 incorporate repeated measures and/or child- and parent-report data).

Figure 2

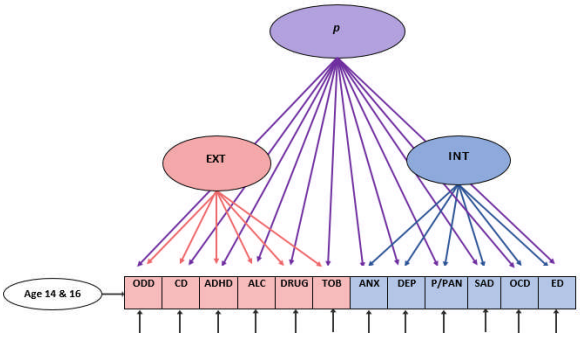
Best-fitting or endorsed CFA measurement model

<u>Child/adolescent population</u>

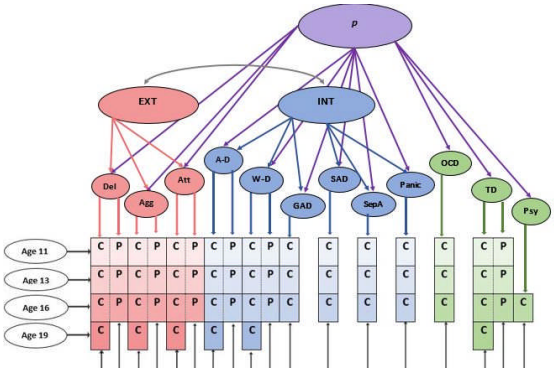
CARRAGHER2016



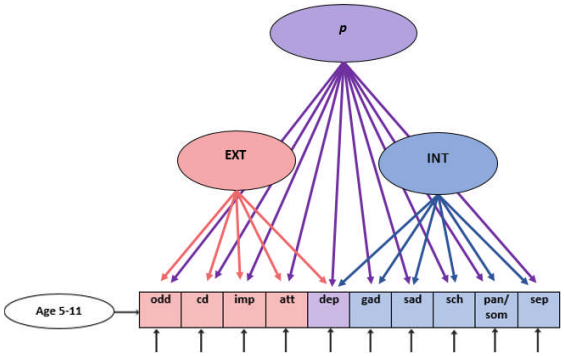
CASTELLANOS2016



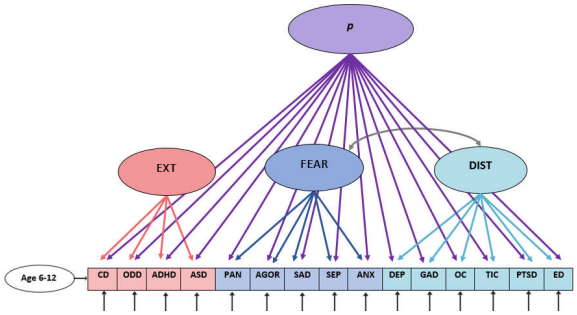
LACEULLE2015



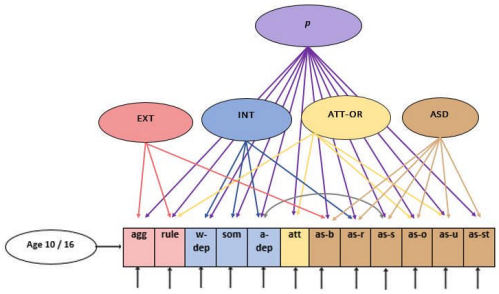
LAHEY2015



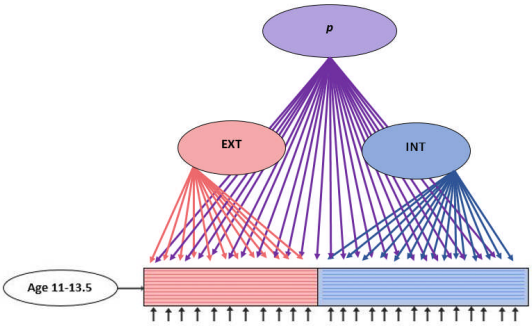
MARTEL2017



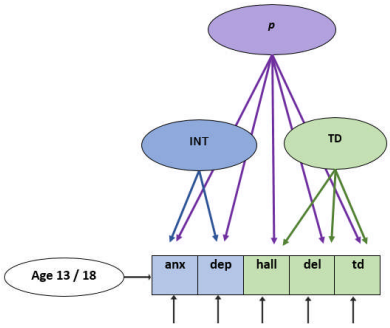
NOORDHOF2015



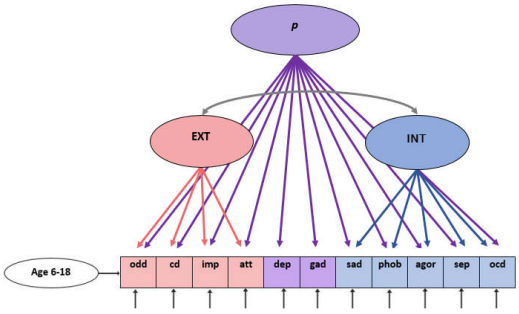
PATALAY2015



STOCHL2015



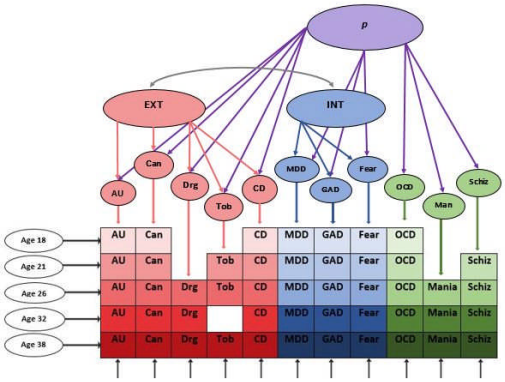
WALDMAN2016



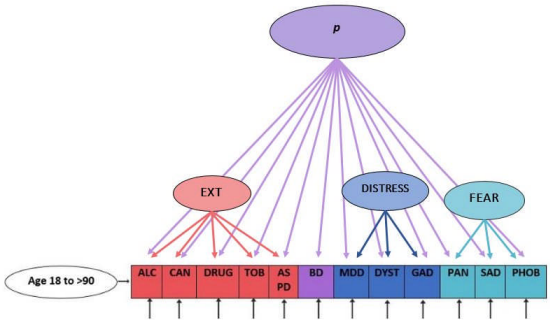
Adult population

Community

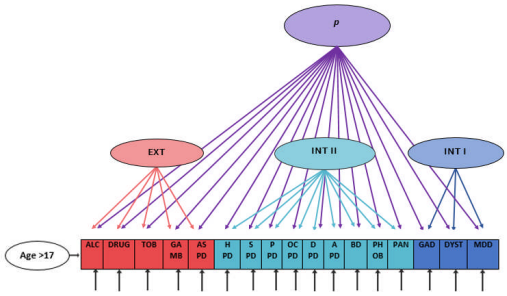
CASPI2014



KIM2015

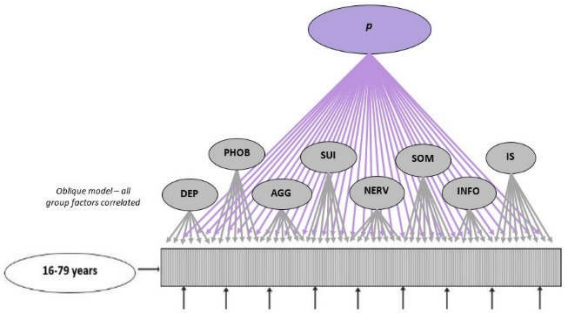


HOERTEL2015

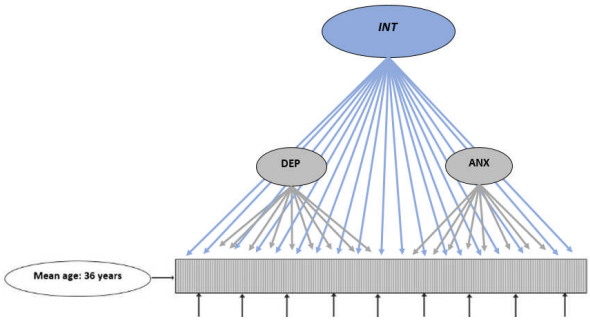


Psychiatric

BRODBECK2014



SUBICA2015



Key:

CFA diagrammatic conventions: Squares/rectangles: observed data; Circles: latent variables; Straight lined, one headed arrow: factor loading/ residual variance; Curved, double headed arrow: factor covariance (oblique models only)

Symptomatology/ data: Lower-case: symptoms (scores); Upper case: disorders (scores); Horizontal stripes: item-level data

Colour code: Purple: p; Pink/red: EXT; Blue: INT; Green: TD; Grey: disorders (where higher-order group factors were not modelled); Others as labelled; Darker colours represent older samples

Characteristics of the p-factor

Relationship between latent factors

As discussed, the motivation for this body of research was the observation that internalising and externalising group factors are correlated in group factor models (Lahey et al., 2012). Several studies specifying oblique bifactor models found that group factors which were correlated in the group factor model either became negatively correlated or had a substantially reduced association in the bifactor version of these models, where *p* was controlled for (Table 7).

Table 7

Comparison of group factor correlations across group and bifactor models, in studies testing oblique models

Group factor model				Bifactor model			
Child/ adolescent population							
CARRAGHER							
	EXT	**INT**	**TD**		**EXT**	**INT**	**TD**
EXT	1			**EXT**	1		
INT	0.449	1		**INT**	0.087	1	
TD	0.474	0.512	1	**TD**	0.287	0.224	1
LACEULLE							
	EXT	**INT**	TD		**EXT**	**INT**	
EXT	1			**EXT**	1		
INT	0.440	1		**INT**	–0.438	1	
TD	0.612	0.883	1				
Adult population							
CASPI2014							
	EXT	**INT**	**TD**		**EXT**	**INT**	
EXT	1			**EXT**	1		
INT	0.328	1		**INT**	-0.471	1	
TD	0.577	0.849	1				

Validity of the *p*-factor

One study considered whether the *p*-factor might reflect a response bias (LAHEY2015) and tested the association between parent-report scales and teacher report of school functioning, global impairment and academic attainment (not shown in Table 5 as these were not used in the CFA measurement model), finding support for the criterion validity of the *p*-factor. KIM2015 derived a hierarchical exploratory factor structure correlated using the Bass-Ackwards method, which they then compared with a bifactor model, finding a high correlation between the Bass-Ackwards unitary factor and bifactor *p* (although the unitary factor would appear to be closer conceptually to a unidimensional than bifactor *p*).

Several of the studies did not investigate the *p*-factor's relationship with variables other than symptomatology at different time points (e.g., LACEULLE2015; KIM2015; NOORDHORF2015). However, the external validity of the *p*-factor was examined by a number of the studies, which tested associations with factors including personality (e.g., CASPI2014; CASTELLANOS2016), cognitive ability or executive function (e.g., CASPI2014; CASTELLANOS2016; LAHEY2015; MARTEL2016), sociodemographic factors (e.g. CASPI2014; PATALAY2015), social competencies and educational attainment (PATALAY2015). WALDMAN2016 was the only study to investigate the genetic heritability of the *p*-factor, however, MARTEL2016 also examined familial risk of psychopathology. The relationship between the *p*-factor and psychopathology (e.g., NOORDHORF2015; PATALAY2015) and suicide risk (HOERTEL2015) at different time points were also investigated. A review of the results of these investigations of the external validity of the *p*-factor would undoubtedly be an important subject for a future review, however, the different methods of analysis used by the studies (for example, SEM extensions to CFA models vs. factor score correlations) and other study-specific differences reviewed above meant that such a review was outside the scope of this study.

Are the p-factor constructs comparable?

In view of these differences in study design and method, a pertinent question is to what degree similar but different models could be said to indicate the same phenomenon. In this area of research, as in mental health research generally, reliable and valid measures are considered to measure a latent variables, and so any differences in how group factors are specified may not be of great conceptual importance. However, *p* represents general covariance across all symptoms measured, which vary considerably across the reviewed studies (Table 5).

It may therefore be helpful to draw a distinction between 'the *p*-factor', a general finding across all the studies reviewed (except SUBICA2015) and a '*p*-statistic', reflecting the study-specific factors outlined above. It should be noted that the fact the *p*-factor (in the general sense) has been extracted across a wide range of symptoms (Table 5) is not equivalent to a *p*-statistic reflecting such a wide symptom range, and this would be a more stringent statistical test of the tendency towards comorbidity implied by this research. For the purposes of referring to a general tendency towards comorbidity, the more pragmatic formulation of 'the *p*-factor' may be appropriate. However, a drawback of this formulation is that it does not offer a way of conceptualizing how broad a transdiagnostic factor would need to be to be considered the *p*-factor; for example, although it may be evident that the general factor modelled by SUBICA2015 is an internalizing factor, meaningful differences between the models tested by NOORDHOF2015 and STOCHL2015 are less easy to describe in these terms. Studies such as KIM2015 which use different modelling methods to extract a hierarchy of interpretable factors imply that these differences are meaningful, and this appears to be a consensus view held by proponents of an empirical nosology (Kotov et al., 2017).

In respect of these formulations of the *p*-factor, it is an open question how far differences between studies are methodological or conceptually important. For example, the decision to artificially keep group factors uncorrelated in orthogonal models inflates shared

variance explained by *p* (Murray, Eisner, & Ribeaud, 2016) and consideration of this methodological decision would be important for accurate comparison of *p*-statistics across studies or populations, however, this may not be of great importance conceptually. Conversely, CARRAGHER2015 suggests that group factors specified by a narrow range of symptoms may not be viable; as CASPI2014 only measured a narrow range of TD symptoms, this has potential implications for the interpretation of their results and therefore the authors' structural hypothesis. Equally, it should be noted that the two studies which found that thought disorder symptomatology did not form a separate group factor (Figure 2) were studies modelling longitudinal data (Table 5). Thought disorders are often enduring and sequentially comorbid (Meyer et al., 2005) and by modelling the additional dimension of duration, the *p*-statistics extracted by these studies may reflect different information about psychopathology. Finally, where studies have not investigated external validity of their *p*-statistics in the same way, the possibility of comparison is limited. For example, LACEULLE2015 directly replicated CASPI2014 in an adolescent population, but as they did not also investigate associated characteristics this leaves open the possibility that comorbidity is influenced by different factors in the different age groups.

Phases 2 and 3: Searching for themes and reporting the thematic analysis

Does the p-factor challenge the diagnostic classification system paradigm and, if so, in what ways?

Diagnoses as discrete categories

Disorders as dichotomous

The most apparent way in which the reviewed studies stand in opposition to the diagnostic paradigm is in their challenge to the ontological assumption that mental health disorders are categorical, dichotomous entities. All the studies were motivated to improve psychiatric nosology and assumed that psychopathology is dimensional, that "diagnostic thresholds increasingly have been acknowledged to be somewhat arbitrary" (CASPI2014; p. 121). This assumption is also evident in the statistical methodology used by all studies, as

latent factors are dimensional even when dichotomous diagnosis data are modelled. However, the reviewed studies retain aspects of the conceptual framework of diagnosis. Several studies modelled diagnosis (Table 5) and the majority of studies discuss their findings in relation to diagnoses.

Disorders as independent of one another

A central premise of this research is that comorbidity is not random, that "mental disorder diagnoses are comorbid at rates much higher than predicted by chance alone" (KIM2015; p. 1064), an assumption borne out by the success of the empirical nosological approach. However, among the reviewed studies there were different positions on the implications of this. The stronger position is that the identification of supra-diagnostic categories challenges the conceptual validity of diagnoses. For example, BRODBECK2014 says "comorbidity among […] disorders challenge[s] categorical classifications of psychopathogical distress" (p. 714) and similar points are made by many of the reviewed studies (discussed below). The weaker position is proposed by KIM2015, who says a "single, optimal comorbidity structure for all purposes is improbable" (p. 1065) and instead suggests that latent factors can be conceptualised within a hierarchical structure, within which disorders start to emerge at particular levels of specificity. According to this weaker view the identification of supra-diagnostic dimensions is not, in itself, a threat to disorders, a view compatible with analogies likening a tendency towards comorbidity to immunodeficiency (PATALAY2015). This weaker position seems more conceptually defensible, given that study authors espousing the stronger position do not consider the (supra-group) *p*-factor to be a threat to the validity of group factors.

The aim of better-understanding the "structure of psychopathology" (comparable phrases were used by almost all studies) through examining comorbidity, represents a shift in the way psychopathology is conceptualised. Within the diagnostic paradigm disorders are dichotomous, and people meeting criteria for diagnoses form an undifferentiated group. Within latent factor research this conception is replaced with an empirical construction of

psychopathology as patterns of covariance at a group level, and, although a particular individual's symptoms could be formulated dimensionally, latent variables themselves are only meaningful at the population level.

Operationalising severity

Within the diagnostic paradigm, disorders are dichotomous and severity is operationalised in different ways, for example as symptom severity or type of disorder ('severe' or 'common'). Rejecting the concept of dichotomous disorders, several of the included studies discussed severity in relation to *p*-factor loadings, with LACEULLE2015 saying that the "[*p*-]factor should be interpreted as characterizing the overall severity of psychopathology" (p.8; similar points were made by CASPI2014 and KIM2015). Several of the studies interpreted the *p*-factor as a dimensional indicator of "general distress" (BRODBECK2014; KIM2015; SUBICA2015). The priority afforded to patterns of comorbidity also impacted on how the reviewed studies conceptualised particular symptoms as indicators of severity (discussed below).

For the above reasons, the reviewed papers did not completely reject the diagnostic paradigm assumption that disorders are discrete categories. However, the remaining assumptions of the diagnostic paradigm (Table 2) will be examined in relation to the broader notion of 'psychopathology', rather than 'diagnoses'.

Psychopathology as latent

Symptoms as observable signs of latent psychopathology

According to the CFA statistical method, "latent continuous factors are hypothesized to account for the pattern of covariance among observed variables" (CASPI2014; p.124). This statistical assumption is expressed through the language used by all the study authors; for example, LACUELLE2015 describes the *p*-factor as "underlying all symptoms of psychopathology" (p. 8) and all studies describe it as a "propensity" or "liability" to psychopathology. For this reason, the criticism of circularity directed at the diagnostic system (Beutler & Malik, 2002) also applies here; the method implies that people experience

comorbidity because they are high-*p,* and they are high-*p* because they experience comorbidity. Operationalising a liability towards comorbidity at the individual level is not straightforward; CARRAGHER2015 acknowledges that an individual could be 'high-*p'* but not receive a diagnosis, and when operationalised with longitudinal data (Table 5) a person's symptoms at any one time may not reflect their 'position' on the dimensional construct of *p*.

Symptoms as independent of each other

Another assumption of CFA is that manifest variables are independent, as otherwise their variation would not indicate a latent factor (Borsboom, 2008). That this might be a simplification of reality and a disadvantage of CFA is acknowledged by NOORDHOF2015, who say the possibility of causally-influential symptoms means "a strong causal interpretation (e.g., problems are directly caused by a [*p*-] factor) of our factor-analytic results may not be warranted" (p. 585). Several other studies acknowledge the possibility of interaction between symptoms, with CASPI2014, LACEULLE2015 and NOORDHOF2015 suggesting dynamic mutualism and BRODBECK2014 suggesting the multiformity model of comorbidity as potential alternative explanations for their results. However, despite these qualifications, the method used by the reviewed studies upholds this assumption and for the most part their results are interpreted accordingly.

Improved symptom characterisations as the route to knowledge of psychopathology

As discussed, within the diagnostic paradigm there is a correspondence relationship (albeit polythetic) between symptoms and the disorder they specify, making an individual's disorder status straightforward to operationalise, and an aim of research is to improve the validity of diagnostic categories. As a body of nosology research, the reviewed studies retained the assumption that improved symptom characterisation is the route to reliable knowledge of psychopathology. Within the empirical nosology the emphasis shifts towards the broad characterisation of the "structure of psychopathology" (as discussed above); and, in the case of several of the studies reviewed here, the status of particular symptoms. For example, LACEULLE2015 is one of several studies which found that INT symptoms loaded heavily on

p, commenting, “this may […] reflect that internalizing problems are more pathological” (p. 8). CASPI2014 and LACEULLE2015 found that TD symptomatology loads directly onto *p* rather than forming a separate group factor (Figure 2), leading them to suggest that “thought problems appear to be core symptoms [of *p*]” (LACEULLE2015; p. 8), an observation which underpins CASPI2014’s ‘structural hypothesis’ of psychopathology. A stronger position is suggested by STOCHL2015, that the *p*-factor implies psychotic phenomena are a more severe rather than a “qualitatively different” presentation, thus apparently employing the empirical approach to suggest qualitative differences to our conceptualization of psychopathology.

Realism about psychopathology

Several of the study authors discuss the “natural classification” (CARRAGHER2016; p. 1) and “structure of psychopathology” (a phrase used by all study authors, as discussed above), as well as the causal properties of particular latent factors; for example, HOERTEL2015 says “the risk of suicide attempts [is raised] through a broad general psychopathological liability” (p. 725). LACEULLE2015 argues the *p*-factor should not be “reified” and several others advise caution in interpreting their results, however, only KIM2015 goes on to discuss this further, saying “reification of a particular factor level […] obscures potentially important variation and construct differentiation across levels of the hierarchy” (p. 1065).

A pragmatic realism is implied by the use of statistics, which is based on the assumption, central to the diagnostic paradigm, that there are entities that exist between people. The reviewed studies appear to go further and imply a more strongly realist attitude; eschewing the idea of dichotomous categories that apply (wholly) to individual, nonetheless hypothesised latent entities are similarly treated in a realist way, although they are instantiated at the population level. Alternatively put, rather than considering statistics a tool given the presumed existence of between-person entities, here the extraction of a statistical construct appears to be considered evidence of its existence.

Key characteristics of psychopathology

Aetiology and the biological basis of psychopathology

An explicit motivating factor for a number of the reviewed studies was that attempts by researchers working within the diagnostic paradigm to identify the aetiology and biological basis of specific disorders had not proved fruitful, but that this situation could be helped by more appropriate clinical groupings (BRODBECK2014; CASPI2014; CASTELLANOS2016; HOERTEL2015; KIM2015; LACEULLE2015; MARTEL2016; PATALAY2015; SUBICA2015). For example, CASPI2014 suggests "researchers should not expect to routinely find single-disorder loyalty in biomarkers […] or causes" (p. 134) as all of the risk factors they tested were primarily associated with *p*, and BRODBECK2014 says the *p*-factor could be an alternative "phenotypic constructs for aetiological research" (p. 725).

Treatment choice and outcomes

Similarly, the studies tend to share the assumption that improved clinical groupings will be associated with treatment outcomes, with all the studies making some mention of clinical implications. Several studies suggest that assessments and interventions should target the "transdiagnostic factor" *p* (BRODBECK2014; CARRAGHER2016; CASTELLANOS2016; HOERTEL2015; KIM2015; SUBICA2015); for example, BRODBECK2014 suggests targeting "underlying liabilities" to develop "interventions that target shared aspects of specific disorders" (p. 12). For the most part, these studies did not offer a more substantive account of how such treatments might work, although two studies (CASPI2014; SUBICA2015) suggest their results support Barlow's unified protocol approach (Barlow et al., 2010), which aims to treat transdiagnostic latent variables by distilling common principles of CBT. Relatedly, several studies make suggestions for treatment based on correlates of *p*, such as personality and traits (CASTALLANOS2016) and symptoms which load heavily onto *p*. For example, CASPI2014 endorses Cognitive Behavioural Therapy (CBT) as *p* is associated with disordered thought and unusual beliefs are ubiquitous in mental health presentations. The unified protocol approach was developed for affective

disorders and perhaps because they also investigate the influence of *p* on ASD and ATT-OR symptoms, NOORDHOF2015, offer a different perspective, saying that the *p*-factor does not imply that "making fine-grained distinction would be unnecessary in clinical practice. For clinical populations […] it can be expected that specific factors become even more important relative to broad factors" (p. 584). However, they do not go further in discussing the relationship between different factors and treatment.

The suggestion that interventions should target the transdiagnostic factors causing symptoms is in line with the latent entity and realism assumptions discussed above. However, demonstrating the efficacy of any treatment on the basis of the evidence of the reviewed studies would not be possible, because no interventions are evaluated. In addition, this assumption begs the question because any statistical model of symptoms will underdetermine a theory about what causes them. The reviewed studies have not identified causal mechanisms and, as discussed above, an explanation of how transdiagnostic factors are associated with symptoms will be circular.

Discussion

Phase 4: Integrative discussion of the results

How far does research into the p-factor challenge the diagnostic paradigm?

Kuhn observed that paradigms are broader than a single theory and constitute a set of assumptions which determine the focus of scientists, as well as informing their beliefs about what entities exist and how knowledge of these entities can be gained (Kuhn, 1996). This thematic analysis has aimed to identify some of the assumptions inherent in the statistical method used by this body of research, as well as its explicit and implicit epistemological and ontological assumptions. The papers reviewed here indicate that in spite of the potentially radical challenge to the diagnostic paradigm which the *p*-factor represents, some key paradigmatic assumptions and beliefs have been retained. In particular, the reviewed studies continued to view psychopathology in terms of latent entities

which cause symptoms and retained a realist attitude towards these entities, leaving them open to the criticism levied against the diagnostic paradigm of circularity. The reviewed studies also retained the assumption that improved classification of psychopathological constructs is to be achieved through reliable specification of symptoms, and several studies reconceptualised qualitative characteristics of particular symptoms on the basis of their tendency to co-vary. Notwithstanding the fact that transdiagnostic constructs are better-defined empirically than diagnoses, as latent factors their explanatory role in any theory may nonetheless remain limited.

What is the p-factor?

The reviewed studies have described a pattern – that there is a spectrum reflecting how far people experience concurrent and persistent comorbidity – which is important information about psychopathology that is lost within the diagnostic paradigm. An outstanding question is how the *p*-factor should be interpreted and what its significance is. This review has aimed to identify unwarranted assumptions which might influence how the *p*-factor is interpreted. It has emphasised that this kind of statistical research will inevitably underdetermine a theory, as alternative explanations could be compatible with the results (Ladyman, 2001). With this in mind, two tentative ways forward for conceptualising the *p*-factor are suggested.

Firstly, statistical studies of the kind reviewed here are unique in describing relationships between factors, including between hypothesised statistical constructs such as the *p*-factor and directly measurable variables such as risk factors. Rather than reifying the *p*-factor by aiming to model the 'structure of psychopathology', contextualising it in terms of associated characteristics could be highly informative. This could involve investigating risk factors, which might differ across different populations and levels of specificity in a hierarchical conception of transdiagnostic factors (KIM2015; Kotov et al., 2017). An extension to this approach could be manipulating models to learn more about how these risk factors operate and to generate testable hypotheses for future research (e.g., CASPI2014;

PATALAY2015). Again, the analogy with heredity may again be instructive, as over several decades population genetics has developed a conceptually rich picture of the way in which environmental and genetic factors interrelate (Haworth & Plomin, 2010). This descriptive approach would be in line with the formulation of '*p*-statistics' outlined in phase 1 of the review.

Secondly, the *p*-factor could be interpreted theoretically; that is, in terms of psychological mechanisms which might explain a general spectrum of comorbidity. The theory of dynamic mutualism, which underpins the network method (Borsboom, 2008), is considered a possible explanation of the *p*-factor by several of the reviewed studies and has been mooted as supporting a possible paradigm shift (McNally et al., 2015). This approach rejects the assumption of the diagnostic paradigm that symptoms are independent and are caused by latent entities; however, dynamic mutualism is focused on symptoms and it is unclear whether it could constitute the type of 'articulated alternative' necessary for a paradigm shift. Theoretical accounts of the *p*-factor have also been offered, including that might reflect emotional and behavioural dysregulation (Beauchaine & Thayer, 2015), the interaction between two dimensions of fast and slow life strategies with neurological integrity predicted by the 'life history theory' (Del Giudice, 2015) and that it reflects a lack of openness to social learning (or 'epistemic trust') in the context of mentalising impairments (Fonagy & Campbell, 2015).

These two ways of thinking about the *p*-factor might not be incompatible. Approached from the perspective of nosology, it might motivate testable hypotheses and theories, such as CASPI2014's structural hypothesis, which raises the question of why persistent affective presentations might be associated with disordered thought; it might also support existing theories emphasising general effects of risk factors leading to latent vulnerability (e.g., McCrory & Viding, 2015) and motivate a fresh perspective on the specificity of other risk factors and psychological processes. As discussed, statistical studies of the kind reviewed

here necessarily underdetermine any theory, however, a theory of psychopathology should account for the pattern of comorbidity identified by the *p*-factor studies.

What are the clinical implications of the p-factor?

The interpretation of the *p*-factor is not merely an academic point, as it could have significant clinical implications. This review urges caution in directly extrapolating from the extraction of statistical constructs to treatment implications, such as unified protocols for psychological therapy. However, building on the above tentative interpretations of the *p*-factor, suggested areas for future consideration are made.

As discussed, studies investigating '*p*-statistics' could be informative for investigating the extent to which risk factors exert a general effect. Interestingly, the issue of prevention is not considered by most of the studies, however, these findings have important implications for targeting environmental risk factors, such as child maltreatment. On the other hand, theoretical accounts of the *p*-factor might have very different clinical implications. Psychopharmacological interventions have been tentatively explored in relation to the dysregulation hypothesis (Beauchaine, 2015), however, of the theoretical accounts of the *p*-factor offered, the implications for psychological intervention have been best-articulated in relation to the epistemic trust hypothesis, although these require empirical validation. Epistemic trust is hypothesised to develop within the context of a mentalising attachment relationship, and the evidence-based model of Mentalization Based Treatment (MBT; Bateman & Fonagy, 2013), has recently been extended to address epistemic trust (Fonagy, Campbell, & Bateman, 2017).

Appraisal of the review

This review aimed to answer interpretive questions about a set of empirical papers and therefore an adapted form of thematic analysis (Braun & Clarke, 2006) was developed. The premise of this method was that the scientific process involves implicit empirical and interpretive methodological assumptions, and that consideration of questions relating to the scientific method requires an integrative critical analysis of both quantitative and qualitative

data. In recent years there have been developments in electronic methods of integrative systematic reviewing, such as text mining (O'Mara-Eves, Thomas, McNaught, Miwa, & Ananiadou, 2015). However, as empirical methods become increasingly complex, they imply conceptual assumptions, and so interpretive approaches to reviewing quantitative research may be increasingly important. This review suggests that implicit conceptual assumptions are important for clinical, as well as academic, reasons and consideration of these issues would therefore seem to be a strength of this review.

However, this review also had limitations. The demands of synthesising quantitative and qualitative data put limits on the breadth and depth of the analysis. It was not possible to examine differences between the child, adolescent and adult studies, or how the studies investigated gender, despite good reasons for thinking age and gender are important factors in this area. It was also not possible to investigate patterns of factor loadings or the external validity of the *p*-factor, although these would be important areas for a future review (MARTEL2016). An additional limitation was the inclusion criteria for the review. In order to interrogate the paradigm, studies which used different methods were excluded (**Figure 1**). However, personality disorder symptomatology was only included in one study, perhaps because recent changes to personality disorder criteria (American Psychiatric Association, 2013) have stimulated more exploratory work. However, bifactor models have been found to fit personality disorder symptomatology, with two recent studies finding that borderline symptomatology loads heavily on the personality disorder general factor (Sharp et al., 2015; Wright, Hopwood, Skodol, & Morey, 2016). Therefore, how the *p*-factor and the general personality disorder factor relate to one other remains and open and pertinent question.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.

Barlow, D. H., Farchione, T. J., Fairholme, C. P., Ellard, K. K., Boisseau, C. L., Allen, L. B., & May, J. T. E. (2010). *Unified Protocol for Transdiagnostic Treatment of Emotional Disorders Therapist Guide* (1 edition). New York: Oxford University Press, USA.

Bateman, A., & Fonagy, P. (2013). Impact of clinical severity on outcomes of mentalisation-based treatment for borderline personality disorder. *The British Journal of Psychiatry*, *203*(3), 221–227.

Beauchaine, T. P. (2015). Future Directions in Emotion Dysregulation and Youth Psychopathology. *Journal of Clinical Child & Adolescent Psychology*, *44*(5), 875–896. https://doi.org/10.1080/15374416.2015.1038827

Beauchaine, T. P., & Thayer, J. F. (2015). Heart rate variability as a transdiagnostic biomarker of psychopathology. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *98*(2 Pt 2), 338–350. https://doi.org/10.1016/j.ijpsycho.2015.08.004

Bentall, R. P. (2004). *Madness explained: Psychosis and human nature*. Penguin UK.

Beutler, L. E., & Malik, M. L. (2002). *Rethinking the DSM: A Psychological Perspective*. American Psychological Association.

Blanco, C., Krueger, R. F., Hasin, D. S., Liu, S.-M., Wang, S., Kerridge, B. T., … Olfson, M. (2013). Mapping common psychiatric disorders: structure and predictive validity in the national epidemiologic survey on alcohol and related conditions. *JAMA Psychiatry*, *70*(2), 199–207.

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, *64*(9), 1089–1108. https://doi.org/10.1002/jclp.20503

Borsboom, D., & Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91–121.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brodbeck, J., Stulz, N., Itten, S., Regli, D., Znoj, H., & Caspar, F. (2014). The structure of psychopathological symptoms and the associations with DSM-diagnoses in treatment seeking individuals. *Comprehensive Psychiatry*, *55*(3), 714–726.

Caron, C., & Rutter, M. (1991). Comorbidity in child psychopathology: Concepts, issues and research strategies. *Journal of Child Psychology and Psychiatry*, *32*(7), 1063–1080.

Carragher, N., Teesson, M., Sunderland, M., Newton, N. C., Krueger, R. F., Conrod, P. J., … Slade, T. (2016). The structure of adolescent psychopathology: a symptom-level analysis. *Psychological Medicine*, *46*(05), 981–994.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., … Poulton, R. (2014). The p factor one general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, *2*(2), 119–137.

Castellanos-Ryan, N., Brière, F. N., O'Leary-Barrett, M., Banaschewski, T., Bokde, A., Bromberg, U., … Gallinat, J. (2016). The structure of psychopathology in adolescence and its common personality and cognitive correlates. *Journal of Abnormal Psychology*, *125*(8), 1039.

Craddock, N., & Owen, M. J. (2010). The Kraepelinian dichotomy–going, going... but still not gone. *The British Journal of Psychiatry*, *196*(2), 92–95.

Del Giudice, M. (2015). The Life History Model of Psychopathology Explains the Structure of Psychiatric Disorders and the Emergence of the p Factor. *Clinical Psychological Science*, *4*(2), 299–311. https://doi.org/10.1177/2167702615583628

Essex, C., & Smythe, W. E. (1999). Between Numbers and Notions A Critique of Psychological Measurement. *Theory & Psychology*, *9*(6), 739–767.

Fonagy, P., & Campbell, C. (2015). Bad Blood Revisited: Attachment and Psychoanalysis, 2015. *British Journal of Psychotherapy*, *31*(2), 229–250. https://doi.org/10.1111/bjp.12150

Fonagy, P., Campbell, C., & Bateman, A. (2017). Mentalizing, Attachment, and Epistemic Trust in Group Therapy. *International Journal of Group Psychotherapy*, *67*(2), 176–201. https://doi.org/10.1080/00207284.2016.1263156

Fuller, S. (2000). *Thomas Kuhn: A Philosophical History for Our Times*. University of Chicago Press.

Geiser, C. (2013). *Data Analysis with Mplus* (1 Tra edition). New York: Guilford Press.

Goldberg, L. R. (2006). Doing it all bass-ackwards: The development of hierarchical factor structures from the top down. *Journal of Research in Personality*, *40*(4), 347–358.

Haworth, C. M. A., & Plomin, R. (2010). Quantitative genetics in the era of molecular genetics: learning abilities and disabilities as an example. *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*(8), 783–793. https://doi.org/10.1016/j.jaac.2010.01.026

Hoertel, N., Franco, S., Wall, M. M., Oquendo, M. A., Kerridge, B. T., Limosin, F., & Blanco, C. (2015). Mental disorders and risk of suicide attempt: a national prospective study. *Molecular Psychiatry*, *20*(6), 718–726.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Kozak, M., Pine, D. S., … Wang, P. (2010). Research Domain Criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiat*, *167*, 748–751.

Jablensky, A. (2016). Psychiatric classifications: validity and utility. *World Psychiatry*, *15*(1), 26–31.

Kendler, K. S. (2016). The nature of psychiatric disorders. *World Psychiatry*, *15*(1), 5–12.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, Severity, and Comorbidity of Twelve-month DSM-IV Disorders in the National Comorbidity Survey Replication (NCS-R). *Archives of General Psychiatry*, *62*(6), 617–627. https://doi.org/10.1001/archpsyc.62.6.617

Kim, H., & Eaton, N. R. (2015). The hierarchical structure of common mental disorders: Connecting multiple levels of comorbidity, bifactor models, and predictive validity. *Journal of Abnormal Psychology*, *124*(4), 1064.

Kinderman, P., Read, J., Moncrieff, J., & Bentall, R. P. (2013). Drop the language of disorder. *Evidence Based Mental Health*, *16*(1), 2–3.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., … Clark, L. A. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454.

Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology*, *2*, 111.

Krueger, R. F., & Markon, K. E. (2011). A dimensional-spectrum model of psychopathology: progress and opportunities. *Archives of General Psychiatry*, *68*(1), 10–11.

Kuhn, T. S. (1996). *The Structure of Scientific Revolutions* (New ed of 3 Revised ed edition). Chicago, IL: University of Chicago Press.

Laceulle, O. M., Vollebergh, W. A., & Ormel, J. (2015). The Structure of Psychopathology in Adolescence Replication of a General Psychopathology Factor in the TRAILS Study. *Clinical Psychological Science*, 2167702614560750.

Ladyman, J. (2001). *Understanding Philosophy of Science* (1st Edition edition). London ; New York: Routledge.

Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is There a General Factor of Prevalent Psychopathology during Adulthood? *Journal of Abnormal Psychology*, *121*(4), 971–977. https://doi.org/10.1037/a0028355

Lahey, B. B., Rathouz, P. J., Keenan, K., Stepp, S. D., Loeber, R., & Hipwell, A. E. (2015). Criterion validity of the general factor of psychopathology in a prospective study of girls. *Journal of Child Psychology and Psychiatry*, *56*(4), 415–422.

Mansolf, M., & Reise, S. P. (2017). When and why the second-order and bifactor models are distinguishable. *Intelligence*, *61*, 120–129. https://doi.org/10.1016/j.intell.2017.01.012

Mareschal, D. (2007). *Neuroconstructivism: How the brain constructs cognition* (Vol. 1). Oxford University Press.

Martel, M. M., Pan, P. M., Hoffmann, M. S., Gadelha, A., do Rosário, M. C., Mari, J. J., … Bressan, R. A. (2017). A general psychopathology factor (P factor) in children:

Structural model analysis and external validation through familial risk and child global executive function. *Journal of Abnormal Psychology*, *126*(1), 137.

McCrory, E. J., & Viding, E. (2015). The theory of latent vulnerability: Reconceptualizing the link between childhood maltreatment and psychiatric disorder. *Development and Psychopathology*, *27*(02), 493–505.

McNally, R. J., Robinaugh, D. J., Wu, G. W., Wang, L., Deserno, M. K., & Borsboom, D. (2015). Mental disorders as causal systems a network approach to posttraumatic stress disorder. *Clinical Psychological Science*, *3*(6), 836–849.

Meyer, S. E., Bearden, C. E., Lux, S. R., Gordon, J. L., Johnson, J. K., O'Brien, M. P., … Cannon, T. D. (2005). The psychosis prodrome in adolescent patients viewed through the lens of DSM-IV. *Journal of Child & Adolescent Psychopharmacology*, *15*(3), 434–451.

Murray, A. L., Eisner, M., & Ribeaud, D. (2016). The Development of the General Factor of Psychopathology 'p Factor'Through Childhood and Adolescence. *Journal of Abnormal Child Psychology*, 1–14.

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies* (Vol. 11). sage.

Noordhof, A., Krueger, R. F., Ormel, J., Oldehinkel, A. J., & Hartman, C. A. (2015). Integrating autism-related symptoms into the dimensional internalizing and externalizing model of psychopathology. The TRAILS Study. *Journal of Abnormal Child Psychology*, *43*(3), 577–587.

O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic Reviews*, *4*(1), 5.

Patalay, P., Fonagy, P., Deighton, J., Belsky, J., Vostanis, P., & Wolpert, M. (2015). A general psychopathology factor in early adolescence. *The British Journal of Psychiatry*, *207*(1), 15. https://doi.org/10.1192/bjp.bp.114.149591

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderheiser, J. (2013). *Behavioral genetics*. Palgrave Macmillan.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696.

Reuters, T. (2016). Endnote (Version X7.7.1).

Sharp, C., Wright, A. G., Fowler, J. C., Frueh, B. C., Allen, J. G., Oldham, J., & Clark, L. A. (2015). The structure of personality pathology: Both general (‘g’) and specific (‘s’) factors? *Journal of Abnormal Psychology*, *124*(2), 387.

Stochl, J., Khandaker, G. M., Lewis, G., Perez, J., Goodyer, I. M., Zammit, S., … Jones, P. B. (2015). Mood, anxiety and psychotic phenomena measure a common psychopathological factor. *Psychological Medicine*, *45*(07), 1483–1493.

Sturt, E. (1981). Hierarchical patterns in the distribution of psychiatric symptoms. *Psychological Medicine*, *11*(04), 783–794.

Subica, A. M., Allen, J. G., Frueh, B. C., Elhai, J. D., & Fowler, J. C. (2015). Disentangling depression and anxiety in relation to neuroticism, extraversion, suicide, and self-harm among adult psychiatric inpatients with serious mental illness. *British Journal of Clinical Psychology*.

Waldman, I. D., Poore, H. E., van Hulle, C., Rathouz, P. J., & Lahey, B. B. (2016). External validity of a hierarchical dimensional model of child and adolescent psychopathology: Tests using confirmatory factor analyses and multivariate behavior genetic analyses. *Journal of Abnormal Psychology*, *125*(8), 1053.

Wolf, E. J., Miller, M. W., & Brown, T. A. (2011). The structure of personality disorders in individuals with posttraumatic stress disorder. *Personality Disorders: Theory, Research, and Treatment*, *2*(4), 261.

Wright, A. G., Hopwood, C. J., Skodol, A. E., & Morey, L. C. (2016). Longitudinal validation of general and specific structural features of personality pathology. *Journal of Abnormal Psychology*, *125*(8), 1120.

Part 2: Empirical Paper

Modelling Axis I and personality disorder symptomatology and its associations with childhood trauma and reflective function

Abstract

Aims: This cross-sectional study aimed to compare alternative models of the comorbidity structure of internalising, antisocial, thought disorder and borderline symptoms, and the relationships between psychopathology and two risk factors which exert general effects across symptom range, childhood trauma and reflective function, were also investigated.

Method: Comprehensive self-report data covering a range of symptoms, internalised representations of childhood traumatic experiences and reflective function were collected as part of the Probing Social Exchanges Study. Confirmatory factor analysis was used to test alternative models of symptomatology, which were compared using standard fit indices. Associations with childhood trauma and reflective function were investigated using outputted latent factor scores.

Results: A bifactor model, with four group factors (internalising, antisocial, thought disorder and borderline) and a general (*p*) factor fitted the data best. Replicating previous findings, correlations between group factors were attenuated when *p* was controlled for in the bifactor model. Childhood maltreatment and reflective function were significantly associated with all symptoms, however, associations with group factor scores were attenuated when *p* was controlled for.

Conclusions: Despite the limitations of an empirical nosology derived from cross-sectional symptoms, the findings presented here provide support for investigation of risk factors for psychopathology within a hierarchical empirical framework.

Introduction

Comorbidity and the p-factor

The diagnostic system

Since it was suggested by Emil Kraeplin at the end of the nineteenth century, the diagnostic system has been the principal framework within which psychopathology has been conceptualised. Significant changes have been made to aspects of the framework and its scope, however, important paradigmatic assumptions have remained unchanged. Central to the diagnostic paradigm are the assumptions that disorders are categorical and dichotomous, have specific aetiologies, biological bases and symptom profiles, and as 'disease' entities are unrelated to one another (Bentall, 2004). The diagnostic system has been criticised on a number of fronts, including for the medicalisation of distress, the lack of acknowledgement of dimensional presentations and for the polythetic criteria by which diagnoses are operationalised (Beutler & Malik, 2002). However, the Diagnostic and Statistical Manual of Mental Disorders fifth edition (DSM-5) taskforce summed up a widely-held opinion when they concluded that, despite increasing awareness that disorder 'categories' are more fluid than previously thought, there is currently insufficient evidence to warrant significant changes to the diagnostic system (American Psychiatric Association, 2013).

Modelling Axis I comorbidity

Clinicians and researchers have long noted the 'comorbidity problem', that disorders appear to not be comorbid at random[2] and in recent decades researchers have used factor analysis, a method based on the assumption that patterns of covariation might reflect an underlying latent factor, to exploit observed comorbidity in order to better-understand psychopathology (Wright, 2016). Exploratory factor analytic (EFA) techniques were first used to model

[2] The 'comorbidity problem' is occasionally presented as the puzzle that comorbidity is observed at all. However, this conception is not used here as there is no inconsistency between the medical model and the presence of multiple disease states.

comorbidity of common mental health disorders, identifying two dimensional latent factors interpreted as ‘internalising’ (anxiety and mood disorders) and ‘externalising’ (antisociality and substance misuse; Krueger, 1999), with subsequent studies extracting ‘fear’ and ‘distress’ as sub-factors of internalising (Vollebergh et al., 2001). These findings have been so robustly replicated across child and adult populations, that in recent years confirmatory factor analysis (CFA) has become more commonly used to model these symptoms. In CFA the pattern of loadings on latent factors (as well as other model specifics, depending on the hypotheses of the researcher) are specified in order to compare alternative hypotheses for explaining the pattern of data observed (Geiser, 2013). This method has proliferated and there is now evidence for additional latent group factors, including thought disorder (Eaton, 2015).

The observation that latent group factors were themselves correlated was the motivation for a series of recent studies which have investigated general as well as specific (‘group’) covariance through bifactor CFA modelling (Lahey et al., 2012). These studies have found broadly similar patterns across child and adolescent populations (Carragher et al., 2016; Laceulle et al., 2015), the general adult population (Caspi et al., 2014; Hoertel et al., 2015) and adult psychiatric populations (Brodbeck et al., 2014; Subica et al., 2015). There is therefore a nascent body of literature suggesting that individuals differ in their tendency to experience any and all mental health disorders as comorbid, sequentially and concurrently, with this dimensional general factor being dubbed the ‘the *p*-factor’ (Caspi et al., 2014).

Studies investigating the *p*-factor have not modelled the same symptoms, which limits how far they can be directly compared. However, several general findings have emerged; firstly, two studies (Caspi et al., 2014; Laceulle et al., 2015) have found that thought disorder symptoms did not form a distinct group factor but instead loaded directly onto p. This pattern lead Caspi and colleagues to suggest a ‘structural hypothesis’ of psychopathology, that internalising and externalising presentations represent pathological versions of gendered personality styles, but that high-p individuals who experienced persistent psychopathology

would be more likely to go on to develop thought disorder, irrespective of gender. Secondly, several studies (Brodbeck et al., 2014; Kim & Eaton, 2015; Laceulle et al., 2015) have found that internalising symptoms have high loadings on *p* and one study (Waldman et al., 2016) found that distress symptoms only loaded onto *p* in their best-fitting model.

Personality disorder symptomatology

Factor analytic approaches have also been used to model personality disorder symptomatology, however, this strand of research has tended to be conducted in parallel. Reasons for this may include that Axis I and personality disorders have historically been seen as qualitatively distinct, being classified on different axes of the DSM since the publication of the third edition (Association, Statistics, & Spitzer, 1980), with personality disorder at that time grouped with intellectual disability and viewed as an essentially unchangeable aspect of a person's character. There are also several reasons pertaining to the factor analytic method. Over different editions of the DSM the diagnostic criteria have changed several times. EFA research has not robustly extracted similar interpretable factors (O'Connor, 2005), unlike models of common mental health disorders. And population studies, which are most commonly used in modelling, have tended not to measure personality disorder symptoms. However, despite these challenges, factor analytic research into personality pathology has recently had a resurgence, having played an important role in the development of the new DSM-5 model for further research (American Psychiatric Association, 2013; Section III) and there is a growing consensus in factor analytic research that personality pathology symptoms can be conceptualised in terms of maladaptive personality traits (Krueger & Markon, 2014).

Perhaps partly due to the fact that personality disorders have tended to be conceptualised within the framework of personality, bifactor EFA modelling techniques were used to interrogate their comorbidity structure before bifactor modelling of Axis I disorders became popular (Jahng et al., 2011; Wolf, Miller, & Brown, 2011b). Several studies have investigated whether borderline personality disorder remains a robust factor in the context of

other personality disorders, finding that borderline symptoms loaded onto a general factor without retaining a group factor (Sharp et al., 2015; A. G. Wright et al., 2016). This has been interpreted as empirical evidence for Kernberg's theory of personality structure, that borderline symptoms are a marker of severity of personality dysfunction (Wright, 2016). These findings have prompted several researchers to consider the relationship between this general personality disorder factor and the Axis I *p*-factor (Sharp et al., 2015; A. G. Wright et al., 2016). However, there has been a paucity of studies investigating the bifactor structure of symptomatology across Axis I and personality disorders; with one such study, crucially, not including borderline symptoms (Hoertel et al., 2015).

Childhood maltreatment, mentalising and the p-factor

A pressing question is how the *p*-factor should be interpreted and to this end several studies have investigated its external validity. Within the diagnostic paradigm, aetiological risk factors are considered to be specific to particular disorders (Beutler & Malik, 2002). However, child trauma and maltreatment have been well-established as risk factors for a number of disorders (Scott, Smith, & Ellis, 2010), including severe presentations such as psychosis (Varese et al., 2012) and personality disorders (Grover et al., 2007), and associated with poorer prognoses and treatment outcomes (Nanni, Uher, & Danese, 2012). One account which might explain the lack of specificity in the effect of child maltreatment is the theory of latent vulnerability, which suggests child maltreatment leads to changes in neurobiological systems which may even be adaptive in the context of abuse or neglect, but which can leave individuals vulnerable to later stressors (McCrory & Viding, 2015). Caspi and colleagues (2014) found that child maltreatment was significantly correlated with p; and that when *p* was controlled for in their bifactor model, significant correlations between the group factors and child maltreatment which had been observed became negligible. However, this has not been investigated further within the *p*-factor literature and so this finding requires replication.

Mentalising or ‘reflective function’, an imaginative capacity to interpret one’s self and others in terms of intentional mental states, develops in the context of an attachment relationship. Two broad mentalising deficits which have been the focus of theoretical and empirical work are *hypomentalising*, characterised by ‘psychic equivalence’, where mental states are experienced concretely, and *hypermentalising* or ‘pretend mode’, characterised by excessive pseudomentalising which may not relate to reality (Bateman & Fonagy, 2012). The ubiquity of social interactions but the lack of predictability of mental states can leave individuals with mentalising difficulties vulnerable to stressors. Attachment anxiety and impairments in mentalising are also hypothesised to lead to poor stress management and a lack of resilience (Luyten, Van Houdenhove, Lemma, Target, & Fonagy, 2012; Nolte, Guiney, Fonagy, Mayes, & Luyten, 2011). Initially proposed in relation to borderline personality disorder, problems with mentalising have been linked to antisocial personality disorder, eating disorders, depression, trauma, addiction (Bateman & Fonagy, 2012), psychosis (Debbané et al., 2016) and functional somatic presentations (Luyten et al., 2012). Until the recent development of a self-report scale, mentalising could only be measured using the Adult Attachment Interview and Parent Development Interview, which limited the possibility of it being investigated in epidemiological studies or other studies with large samples (Fonagy et al., 2016). In light of evidence linking mentalising to a range of disorders, there is a need to investigate its relationship to the *p*-factor.

This study

This study was a part of a larger research project, the ‘Probing Social Exchanges Study’, within which comprehensive self-report data were collected from individuals diagnosed with a personality disorder and non-clinical control participants. In CFA different models can be considered formalisations of alternative hypotheses to explain the data. Five CFA models of externalising, internalising, thought disorder and borderline symptoms were specified, alongside the unifactorial model typically used as the standard null model in CFA.

- There is a large body of literature indicting that group factor models describe the comorbidity of psychopathology (Krueger & Markon, 2011). Therefore, a correlated group factors model with four groups (externalising, internalising, thought disorder and borderline) was specified.
- Recent studies have indicated that bifactor models, with each indicator variable loading onto a group and general factor, best describe the comorbidity structure of psychopathology (Carragher et al., 2016). Therefore a full bifactor model with four group factors (externalising, internalising, thought disorder and borderline) and a general factor (*p*) was specified.
- Several studies found that thought disorder symptoms did not form a separate factor in a bifactor model structure (Caspi et al., 2014; Laceulle et al., 2015). Therefore a modified bifactor model with three group factors (externalising, internalising and borderline) and a *p*-factor was specified.
- Several investigations of personality disorders have found that borderline symptomatology does not form a group factor in a bifactor model of personality disorder symptoms (Sharp et al., 2015; Wright et al., 2016). Extrapolating from these findings to this study, which includes Axis I symptoms, a modified bifactor model with three group factors (externalising, internalising and thought disorder) and a *p*-factor was specified.
- Several studies found that internalising symptoms have high loadings on *p* or do not load onto an additional group factor (Brodbeck et al., 2014; Kim & Eaton, 2015; Laceulle et al., 2015; Waldman et al., 2016). Therefore a modified bifactor model with three group factors (externalising, thought disorder and borderline) and a *p*-factor was specified.

Finally, given the lack of specificity in the relationships between psychopathology and both childhood maltreatment and mentalising, it was hypothesised that these would be related to latent factors. This hypothesis was tested using the best-fitting CFA model.

Method

Study details and data collection

The Probing Social Exchanges Study

This study is part of a larger project investigating the neural correlates and computational mechanisms of social processes relevant to personality disorder in adolescents and adults, which has been running since June 2012. The study is based at the Wellcome Trust Centre for Neuroimaging, University College London, where all participants were tested, in collaboration with the Virginia Tech Carilion Research Institute, Virginia, United States. NHS ethical approval was granted by the Research Ethics Committee of Wales (12/WA/0283). Communication from the committee and information provided to participants are included in Appendices B2-B5.

The Probing Social Exchanges Study involves the collection of comprehensive cross-sectional assessment data on people diagnosed with borderline or antisocial personality disorders and non-clinical controls, which takes several days to collect for each participant. Trainee Clinical Psychologists are able use data collected from all participants, conditional on joining the study research team, attending training and conducting behavioural tests and interviews on a small number of participants. This particular project was restricted to the adult part of the Probing Social Exchanges Study.

Recruitment and inclusion criteria

Participants with borderline personality disorder were recruited from 24 clinical services in London, participants with antisocial personality disorder were recruited from three probation services and non-clinical controls were recruited following their responding to advertisement material distributed through various media.

To join the adult part of the study participants needed to be aged between 18-65 years, to be fluent in spoken and written English and to have normal corrected vision. Exclusion criteria were current or past history of neurological disorders or trauma (including

epilepsy, head injury and loss of consciousness) and a learning disability requiring specialist educational support or medical treatment. People with active psychosis were excluded from the study but there were no further exclusion criteria relating to psychopathology for either clinical or control group. There were further exclusion criteria for a neuroimaging component of the study, however, these did not affect the behavioural testing and so were not relevant to the present study.

Sample

The purpose of this study was to reconceptualise psychopathology dimensionally rather than in the terms of the diagnostic paradigm. In line with current thinking (Insel et al., 2010) it was decided that this implied the use of a combined clinical and non-clinical sample. The combined sample was therefore considered a purposive sample, with over-sampling of a particular clinical presentation. The inclusion criteria for the Probing Social Exchanges Study are broad, with no exclusion criteria for psychopathology other than active psychosis, and with personality disorder being the only mental health criterion used to differentiate clinical and control participants. Within the diagnostic paradigm comorbidity is assumed to be random, however, in reality, comorbidity is especially common in people with personality disorders (Skodol et al., 2002) and it was expected that using this combined sample would inflate correlations between all symptoms. The decision to use a combined sample was made on conceptual grounds, however, the proportion of participants meeting the clinical cut-off for borderline personality disorder on the Personality Assessment Inventory – Borderline Features (PAI-BOR) were compared; 85% of the personality disordered participants and 19% of the controls were above the clinical cut-off. It is unclear whether this overlap between the groups demonstrates a weakness of self-report measures (Stone, Bachrach, Jobe, Kurtzman, & Cain, 1999) or whether there was potentially undiagnosed personality pathology in the control group, although a clinical score on the PAI-BOR does not constitute a diagnosis.

521 adult participants were tested by the study team (350 personality disorder; 171 controls). Of these, 16 either withdrew or did not attend all testing days so there was no self-report data available (12 personality disorder; four controls). Demographic characteristics were compared across the two groups of participants, revealing that there were significant differences between them (Table 1). Several studies modelling the *p*-factor have found that, whilst internalising and externalising psychopathology are differentially associated with gender, *p* itself is not; however, gender was not included in the analysis here because significantly more participants with a personality disorder were female.

Table 1

Sample demographic characteristics, by personality disorder and control participants

	Participants		**Difference statistic**	**Combined sample**
	Personality disorder	**Control**		
N	338	167		505
Mean age (sd)	32.3 (10.5)	30.1 (11.0)	*t*(500) = -2.16, *p*= .031	31.6 (10.7)
Age range	18-65	18-62		18-65
Gender				
- Female	232	100	Chi-sq(1) = 4.8, *p*= .028	332
- Male	99	66		165
- Not recorded	7	1		8
Ethnicity				
- White – British, Irish, any other white background	249	101	Chi-sq(4) = 11.3, *p*= .023	350
- Black/ Black British	27	19		46
- Asian or British Asian	20	20		40
- Mixed ethnicity	27	20		47
- Other/not stated	9	7		22

Symptom measures

Symptom data used in the analysis were dimensional scores on self-report measures. As modelling symptom-level data is more accurate (Carragher et al., 2016), where possible, particular symptoms were differentiated by using subscales or validated scale factors.

Antisocial Process Screening Device

The Antisocial Process Screening Device (APSD) is a 20-item measure of psychopathy in adolescents, with two subscales; *Impulsive/conduct problems* and *Callous/unemotional traits*. The APSD has been found to have good internal consistency and external validity in adolescent samples (Munoz & Frick, 2007), however, it has not been validated in adult samples.

Brief Symptom Inventory

The Brief Symptom Inventory (BSI) is a 53-item measure derived from the Symptom Checklist-90. It has nine subscales; *Somatization*, *Obsessive-compulsive*, *Interpersonal sensitivity*, *Depression*, *Anxiety*, *Hostility*, *Phobic anxiety*, *Paranoid ideation* and *Psychoticism*. It is widely-used and has been validated in non-clinical and clinical outpatient and inpatient samples, and has very good test-retest reliability and internal consistency, and good convergent and construct validity (Derogatis & Melisaratos, 1983).

Difficulties in Emotion Regulation

Difficulties in Emotion Regulation (DERS) is a 36-item measure of awareness, understanding and acceptance of emotions, the ability to refrain from impulsive behaviour when experiencing emotion and emotion regulation strategies, aspects of emotional regulation which are negatively associated with borderline personality disorder (Salsman & Linehan, 2012). The scale has high internal consistency, good test–retest reliability and adequate construct and predictive validity (Gratz & Roemer, 2004).

Green et al Paranoid Thoughts Scale

The Green et al Paranoid Thoughts Scale (GPTS) is a 32-item scale measuring ideas of social reference and ideas of persecution. The scale has good internal consistency, test-retest reliability and concurrent and convergent validity (Green et al., 2008).

Life History of Aggression

Life History of Aggression (LHA) is a 10-item measure with three sub-scales; *Aggression*, *Consequences/antisocial behaviour* and *Self-directed aggression* (hereafter referred to as 'self-injury'). The LHA and its sub-scales have excellent test-retest and interrater reliability, internal consistency and concurrent validity (Coccaro, Berman, & Kavoussi, 1997).

Personality Assessment Inventory – Borderline Features

PAI-BOR is a 24-item scale which is part of the Personality Assessment Inventory, a battery of tests covering all aspects of personality (Moray, 1991). The PAI-BOR has four subscales measuring different symptoms of borderline personality; *Affective instability*, *Identity disturbance*, *Negative relationships* and *Self-harm* (hereafter referred to as 'self-defeating behaviour' for clarity, as it does not measure bodily self-injury). The PAI-BOR has good interrater reliability and criterion validity (Stein, Pinsker-Aspen, & Hilsenroth, 2007).

PTSD Checklist – Specific

The PTSD Checklist – Specific (PCL-S) is an 17-item measure of post-traumatic stress disorder (PTSD) which measures the symptoms of re-experiencing, avoidance, dysphoria/numbing and hyperarousal (Weathers, Litz, Herman, Huska, & Keane, 1993). It is the most widely-used measure of PTSD and has been shown to have good test-retest and criterion validity (McDonald & Calhoun, 2010).

Schizotypal Personality Questionnaire

The Schizotypal Personality Questionnaire (SPQ) is a 74-item measure of schizotypal personality disorder, with nine subscales measuring different symptoms of the disorder. The SPQ has been found to have good test-retest reliability and good convergent, discriminant and criterion validity in non-clinical samples (Raine, 1991). Factor analytic research identified three distinct factors measured by the scale; cognitive-perceptual difficulties, interpersonal difficulties and disorganised speech and behaviour (Raine et al., 1994). The four subscales relating to the cognitive-perceptual difficulties factor (*Magical thinking*; *Unusual perceptions*; *Ideas of reference* and *Suspiciousness*) were used in this study.

Symptom grouping factors

A simplifying assumption often made in CFA is that manifest data (in this case, symptoms) only load onto one group factor, with decisions regarding factor specification often being made on the basis of prior exploratory analysis in the literature (Wright, 2016). That CFA does not usually allow cross-loadings means there is 'unmodelled complexity', which inevitably leads to biased parameter estimates and which needs to be taken into account when interpreting findings (Aja L. Murray & Johnson, 2013). A particular issue here is the specification of symptoms which might be related to more than one latent factor. The significance of this may depend on the disorder in question, both with regards actual co-occurrence of symptoms and diagnostic conventions. For example, borderline personality disorder symptoms include anger, paranoia, anxiety and mood disturbance, also symptoms of other disorders, but, conversely, psychotic experiences, which may in reality be a common experience in anxiety and mood disturbance (Stochl et al., 2015), are exclusion criteria for these disorders (American Psychiatric Association, 2013).

A robust finding in the literature (Krueger & Markon, 2011) is that a range of antisocial behaviours load onto an externalising factor (here this factor was named 'ASOC', as substance misuse was not measured). It was noted that the 'hostility' subscale of the BSI was potentially ambiguous as three of the five items ask about angry outbursts and arguments, which could be considered on a borderline spectrum, but it was retained on ASOC as the remaining items ask about urges to do physical damage to people or objects.

A similarly robust finding in the literature is that anxiety and mood disturbance load onto an internalising factor ('INT'; Krueger & Markon, 2011). Caspi and colleagues (2014) modelled obsessive-compulsive behaviour on their thought disorder factor, however, they found similar results when they repeated their analysis with it loading onto internalising, as is a more commonly practice (Miller et al., 2012). Internalising and externalising sub-groups of post-traumatic stress have been identified (Miller 2003; 2004) and the symptom of re-experiencing could be considered to be part of the thought disorder spectrum, however, it

has more commonly been considered an anxiety disorder (Martel et al., 2017) and here it was retained on INT. Somatic presentations have been less commonly included in latent models and they may constitute a separate dimension (Kotov et al., 2011), however, the 'somatising' subscale of the BSI includes items relating to anxiety (Brodbeck et al., 2014) and so this symptom was retained on INT. Finally, interpersonal sensitivity was retained on INT, however, it was noted that this could be considered part of a borderline spectrum of symptoms.

A full range of thought disorder symptoms have not been modelled, although there is evidence of a factor specified by hallucinations and delusions (Kotov et al., 2011; Krueger & Markon, 2011), and so the symptoms measured here were included on a thought disorder ('TD') factor on conceptual grounds. It was noted that paranoia and persecutory thoughts might be considered part of a borderline spectrum. Active psychosis was an exclusion criteria for the study and was not measured.

Finally, several studies have found that borderline personality forms a coherent factor (reviewed in Wright, 2017) and on this basis borderline symptoms were retained on a single factor ('BOR'), although it was noted that studies which have investigated borderline symptoms have tended to model them alongside symptoms of other personality disorders.

To allow model identification, group factors should be specified by at least three manifest variables (Geiser, 2013) and the number of manifest variables should be approximately equal across factors (Reise, 2012); these final checks indicated the group factor structure was viable. Details of symptom measurement, by group factor, are shown in Appendix B6.

Additional measures

Childhood Trauma Questionnaire

The Childhood Trauma Questionnaire (CTQ) is a 28-item retrospective measure of child abuse and neglect. Items are scored on a five-point Likert scale and recoded scores

comprise five subscales; *Physical abuse*, *Sexual abuse*, *Emotional abuse*, *Physical neglect* and *Emotional neglect*, as well as a total score. The CTQ has good test-retest reliability and convergence with the Childhood Trauma Interview, and its subscales have high internal consistency (Bernstein et al., 1994).

Reflective Function Questionnaire

The Reflective Function Questionnaire (RFQ) is a 54-item scale with two subscales measuring the distinct constructs of Certainty about mental states (RFQ_C) and Uncertainty about mental states (RFQ_U). Mentalising is a complex construct which is challenging to measure, as self-report scales require self-knowledge, and so raw scores (on a seven-point Likert scale) are recoded in order to avoid individuals' misperceptions of their mentalising ability cofounding results. Low scores on the RFQ_C are intended to reflect hypermentalising, whilst higher scores reflect more genuine mentalising ability; high scores on the RFQ_U are intended to reflect hypomentalising, whereas lower scores reflect mentalising ability. Both subscales have satisfactory internal consistency and test-retest reliability across clinical and non-clinical samples (Fonagy et al., 2016).

Method of analysis

Required sample size

Estimation of power of a model to derive a test of good fit can be based on the known distribution of the root mean square error approximation (RMSEA) statistic (MacCallum, Browne, & Sugawara, 1996). Given an alpha level of .05, desired power of 0.8 and the degrees of freedom for the least complex model tested (Table 6), the required sample size for a test of close fit was calculated using SAS syntax (SAS Institute, n.d.) provided by MacCallum and colleagues. With the null RMSEA value set at .05 and the alternative RMSEA at .08, the required sample size was estimated to be 74. Although small sample size estimates should treated with caution, where the degrees of freedom are large, moderate to large sample sizes allow for extremely high power to detect close fit (MacCallum et al.,

1996). Therefore the sample size of this study was judged to be adequate for a test of close fit using RMSEA.

In addition to power to assess overall model fit, further considerations in CFA modelling are potential bias in parameter estimates, bias in standard errors and solution propriety, as models tested with small samples may fail to converge without improper solutions. Factors influencing required sample size are the number of latent factors, factor loadings and amount of missing data (Wolf, Harrington, Clark, & Miller, 2013). Guidelines regarding these factors were considered in relation to this study and were judged not to be of concern, confirming the sample size was likely to be adequate.

Pre-analysis data checks

Data were cleaned using SPSS version 24 (IBM, 2016), during which missing data were assessed (Appendix B7). Simple imputation drawing on predictive distributions is a robust method of dealing with missing data, albeit resulting in a lack of precision due to underestimated standard errors (Little & Rubin, 2014). Where five percent or less of item-level data were missing for a particular measure and participant, this was imputed using the SPSS Expectation–Maximisation algorithm, which yields reliable estimates (Enders, 2003).

At the scale or subscale level, cases where particular questionnaires had not been administered were assumed to be missing at random. For the remaining missing symptom scale scores, Little's missing completely at random test indicated that data could be assumed to be missing at random (Chi-sq(1232) = 1308.8, $p = .063$).

Skewness and kurtosis were calculated (Table 2) and the Shapiro-Wilk test of normality indicated that the data were not normally distributed ($p < .001$ for all variables), which was accounted for in the analysis.

Confirmatory factor analysis

Models tested

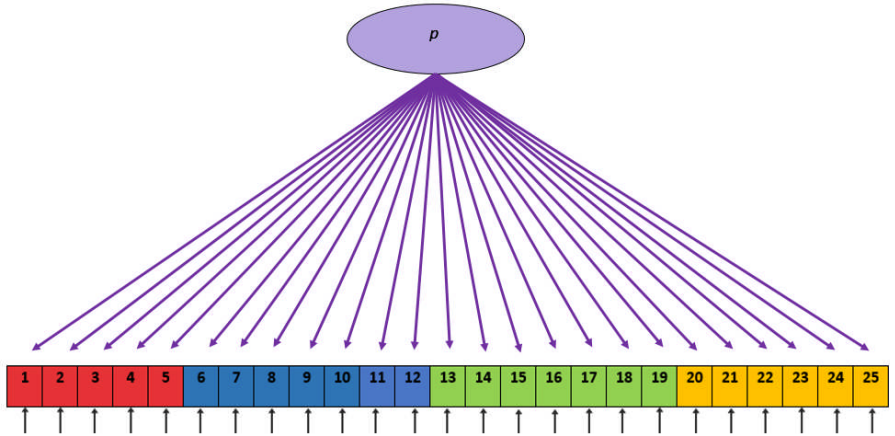
Traditionally bifactor models are orthogonal, which inflates the variance explained by *p* (Martel et al., 2017). However, a strong body of evidence that psychopathology symptoms and group factors are correlated means there is a conceptual argument against artificially setting correlations between factors at zero (Wright, 2017), so all specified models were oblique at the group factor level. An additional benefit of oblique models is that they allow for comparison of the relationships between group factors. With these considerations taken into account, six CFA models were specified (Figure 3).

Modelling software and specification

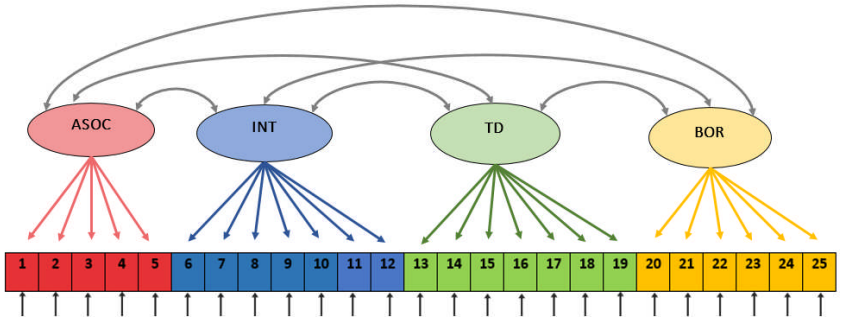
All models were estimated using Mplus version 7 (Muthén & Muthén, 2012). Modelling specification and syntax are shown in Appendices B8 and B9.

Figure 3: CFA models tested

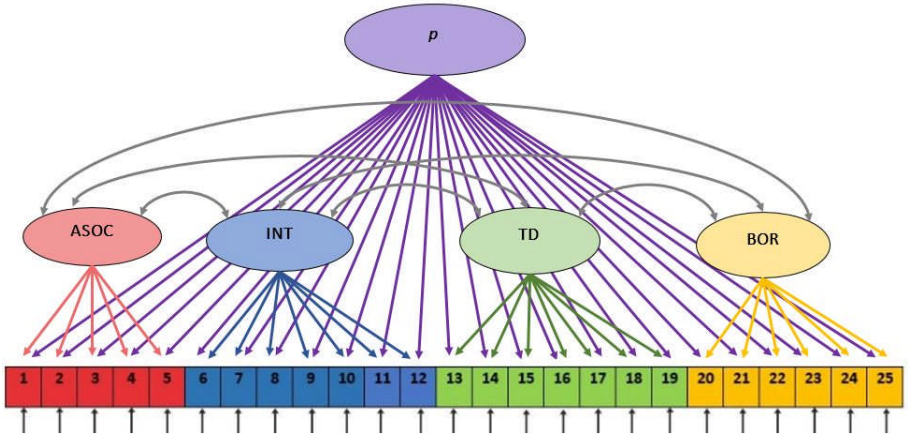
Model 1: Unifactorial



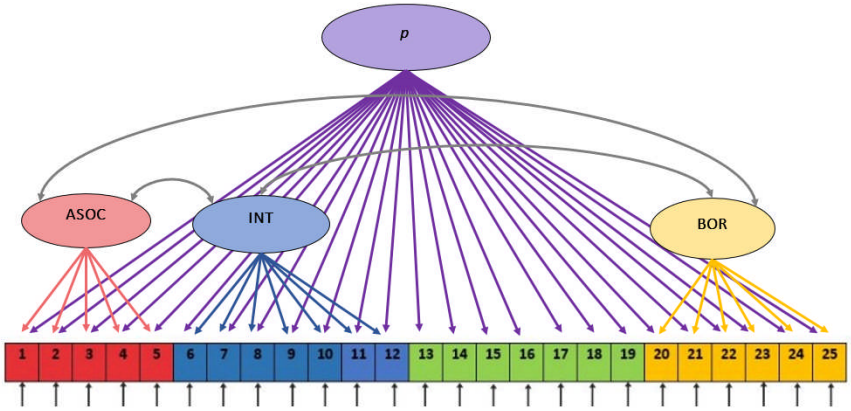
Model 2: Correlated group factors



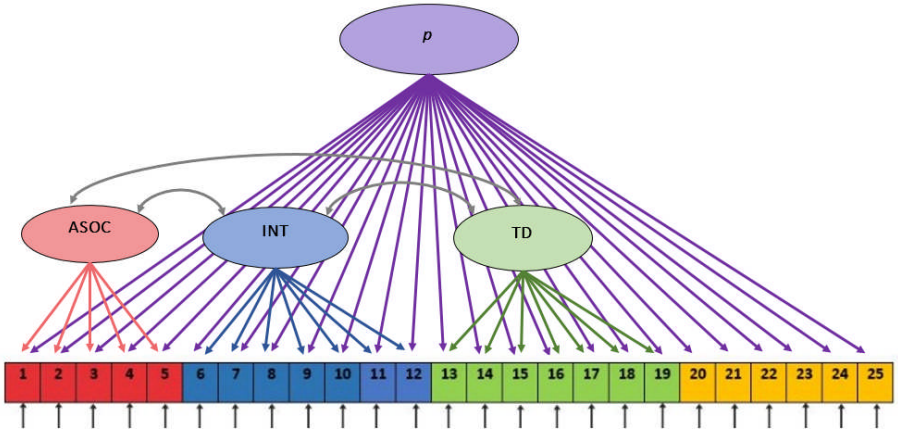
Model 3: Full bifactor



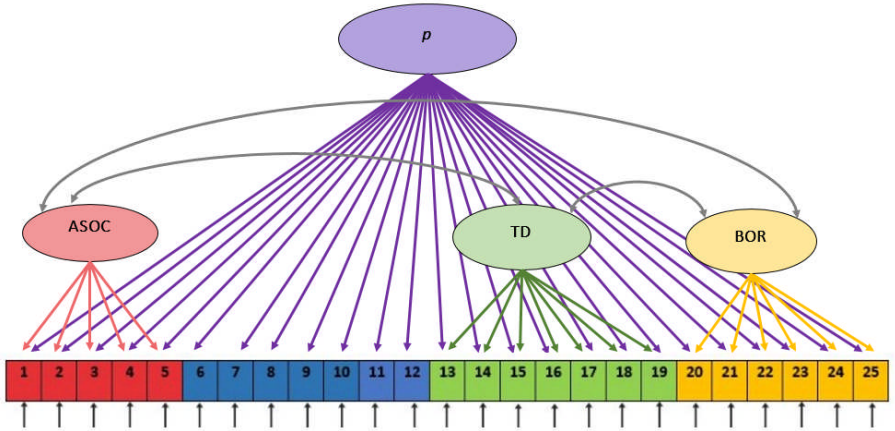
Model 4: Modified bifactor (no TD factor)



Model 5: Modified bifactor (no BOR factor)



Model 6: Modified bifactor (no INT factor)



Key

CFA diagrammatic conventions: Squares/rectangles: observed data; Circles: latent variables; Straight lined, one headed arrow: factor loading/ residual variance; Curved, double headed arrow: factor covariance

Colour code: Purple: *p*; Pink/red: EXT; Blue: INT; Green: TD; Yellow: BOR

Symptoms: 1 – Hostility; 2 – Aggression; 3 – Antisocial behaviour; 4 – Impulsivity; 5 – Callous/ unemotional; 6 – Somatising; 7 – Interpersonal sensitivity; 8 – Depression; 9 – Anxiety; 10 – Specific phobia; 11 – Post-traumatic stress; 12 – Obsessive compulsivity; 13 – Psychoticism; 14 – Magic thinking; 15 – Unusual perceptions; 16 – Suspiciousness; 17 – Ideas of reference; 18 – Paranoid thoughts; 19 – Persecutory thoughts; 20 – Affective instability; 21 – Emotional dysregulation; 22 – Identity problems; 23 – Negative relationships; 24 – Self-defeating; 25 – Self-injury

Model fit indices

Six indicators were used to evaluate the models. RMSEA and standardised root mean square residual (SRMR) are absolute fit indices, with general guidance being that values of 0.06 and 0.08 respectively indicate a good fit (an RMSEA value of 0.08 indicates an adequate fit). The comparative fit index (CFI) and Tucker-Lewis index (TLI) are relative fit indices; values above 0.90 may be considered an adequate fit and 0.95 or more indicates excellent fit (Hu & Bentler, 1999). Finally, the Akaike information criterion (AIC) and sample size-adjusted Bayesian information criterion (ABIC) are indicators of model parsimony, with lower values indicating more parsimonious and preferable, models (Kotov et al., 2011). The Chi-sq value test of model fit is typically significant where sample sizes are large and so was not used, although it is given with the other fit indices in Table 6.

Relationship between symptomatology and other factors

Spearman's rho correlation coefficients of the relationship between each symptom and the subscales of the CTQ, CTQ total score, RFQ_U and RFQ_C were calculated, to check whether the hypothesised lack of specificity of association between these factors and the symptoms were observed.

There are two ways in which the association between latent factors and additional variables can be investigated; using outputted latent factor scores or by extending a CFA model to include the additional factors. Factor scores are not distributed exactly as 'true' factors, which is a drawback of this method (Carragher et al., 2016), however, they allow for the CFA model to be fixed, such that parameter estimates are not influenced by inclusion in the model of additional factors, and so this method was judged appropriate for the study hypotheses. Factor scores are estimated in Mplus using a Bayes estimator (Muthén & Muthén, 2006), these were then saved and Spearman's rho correlation coefficients for the relationships with CTQ, RFQ_U and RFQ_C were calculated in SPSS.

Results

Descriptive statistics

Descriptive statistics for each variable are shown in Table 2.

Table 2

Mean, standard deviation, number of observations and indicators of normality

	Variable	N	Mean	Sd	Skewness	Kurtosis
Symptoms						
ASOC						
1	Hostility	474	6.04	5.56	0.85	-0.38
2	Aggression	490	13.74	6.64	-0.03	-0.90
3	Antisocial behaviour	495	5.32	4.97	0.90	0.17
4	Impulsivity	460	6.85	3.48	0.23	-0.59
5	Callous/ unemotional	461	3.66	2.12	0.66	0.20
INT						
6	Somatising	474	7.65	6.93	0.80	-0.29
7	Interpersonal sensitivity	473	6.60	5.32	0.28	-1.32
8	Depression	473	10.71	8.11	0.15	-1.42
9	Anxiety	473	8.41	7.11	0.49	-0.98
10	Specific phobia	473	5.73	5.86	0.80	-0.58
11	Post-traumatic stress	489	50.06	20.19	-0.13	-1.28
12	Obsessive compulsivity	473	10.89	7.18	0.14	-1.25
TD						
13	Psychoticism	473	6.72	5.53	0.35	-1.03
14	Magic thinking	486	1.86	1.94	0.94	-0.01
15	Unusual perceptions	489	3.54	2.76	0.36	-1.00
16	Suspiciousness	492	4.73	2.90	-0.32	-1.38
17	Ideas of reference	494	4.25	2.84	0.00	-1.17
18	Paranoid thoughts	473	6.99	5.82	0.43	-0.96
19	Persecutory thoughts	467	66.18	34.31	0.92	-0.24
BOR						
20	Affective instability	501	12.16	4.64	-0.45	-0.90
21	Emotional dysregulation	496	113.24	35.97	-0.25	-1.16
22	Identity problems	501	11.48	4.80	-0.39	-0.95
23	Negative relationships	501	11.98	4.34	-0.50	-0.73
24	Self-defeating	501	9.50	5.11	0.09	-1.08
25	Self-injury	497	4.47	3.76	0.07	-1.53
Additional variables						
Childhood physical abuse		465	9.65	6.03	1.18	0.25
Childhood sexual abuse		463	8.75	6.49	1.53	0.94
Childhood emotional abuse		467	14.05	6.86	0.26	-0.99
Childhood physical neglect		466	9.85	4.56	0.95	0.26
Childhood emotional neglect		463	15.10	6.27	-0.12	-1.19
Childhood trauma total (CTQ total)		460	57.32	24.36	0.59	-0.45
Certainty about mental states (RFQ_C)		469	18.32	13.53	0.94	0.57
Uncertainty about mental states (RFQ_U)		469	22.24	15.14	0.75	0.04

Symptom correlations

Spearman's rho correlation coefficients were calculated and are shown in Table 3. Several observations were made on the basis of examining the correlation matrix. Firstly, the data does not support the assumption that comorbidity exists at random, as significant correlations between symptoms were observed in all cases.

Secondly, each of the factors appeared to be coherent; that is, there were significant correlations of at least moderate magnitude between the symptoms within each factor group. These intra-factor correlations were highest for the INT factor (0.669-0.853) and lowest for ASOC (0.309-0.646). In addition, ASOC symptoms tended the have the lowest correlations with symptoms outside of their group factors. Hostility was the only ASOC symptom which had correlations of a moderate magnitude with symptoms of other factors, whereas callous/unemotional had the lowest correlations with other symptoms, both within and outside ASOC. The INT symptoms had very high intra- and inter-factor correlations with other symptoms. As might be expected (Caspi et al., 2014) obsessive-compulsivity was highly correlated with TD symptoms, however, its highest correlations were with other INT symptoms, supporting its inclusion on INT. Similarly, post-traumatic stress correlated highly with INT, and interpersonal-sensitivity correlated more highly with INT than BOR.

Table 3

Spearman's rho correlation coefficients between symptoms, organised by group factor

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1 Hostility	1																								
2 Aggression	.573**	1																							
3 Antisocial behaviour	.427**	.649**	1																						
4 Impulsivity	.596**	.540**	.494**	1																					
5 Callous/ unemotional	.348**	.309**	.330**	.457**	1																				
6 Somatising	.624**	.394**	.290**	.434**	.242**	1																			
7 Int'personal sensitivity	.667**	.374**	.261**	.479**	.183**	.669**	1																		
8 Depression	.683**	.371**	.271**	.469**	.254**	.723**	.835**	1																	
9 Anxiety	.703**	.421**	.301**	.492**	.244**	.792**	.827**	.825**	1																
10 Specific phobia	.647**	.412**	.281**	.496**	.248**	.699**	.811**	.786**	.853**	1															
11 Post-traumatic stress	.649**	.454**	.311**	.510**	.288**	.688**	.709**	.729**	.783**	.745**	1														
12 Obsessive compulsive	.673**	.432**	.310**	.492**	.282**	.730**	.776**	.799**	.823**	.779**	.735**	1													
13 Psychoticism	.716**	.407**	.298**	.495**	.286**	.713**	.833**	.879**	.835**	.806**	.758**	.798**	1												
14 Magic thinking	.414**	.349**	.262**	.244**	.098*	.436**	.336**	.319**	.405**	.378**	.371**	.366**	.367**	1											
15 Unusual perceptions	.569**	.419**	.283**	.438**	.257**	.597**	.510**	.530**	.621**	.578**	.625**	.613**	.610**	.583**	1										
16 Suspiciousness	.670**	.443**	.343**	.541**	.308**	.602**	.703**	.650**	.676**	.700**	.700**	.661**	.708**	.438**	.632**	1									
17 Ideas of reference	.581**	.407**	.305**	.480**	.239**	.570**	.602**	.568**	.613**	.609**	.596**	.615**	.621**	.529**	.684**	.731**	1								
18 Paranoid thoughts	.744**	.474**	.359**	.553**	.305**	.671**	.780**	.760**	.788**	.753**	.712**	.749**	.796**	.424**	.588**	.775**	.639**	1							
19 Persecutory thoughts	.696**	.481**	.414**	.528**	.285**	.644**	.746**	.703**	.746**	.715**	.691**	.695**	.742**	.478**	.627**	.780**	.677**	.844**	1						
20 Affective instability	.719**	.556**	.378**	.563**	.271**	.567**	.662**	.648**	.692**	.673**	.725**	.644**	.648**	.328**	.523**	.651**	.556**	.633**	.616**	1					
21 Emotional dysreg'	.627**	.405**	.248**	.534**	.268**	.619**	.770**	.753**	.746**	.757**	.768**	.715**	.746**	.280**	.551**	.702**	.592**	.657**	.655**	.771**	1				
22 Identity problems	.577**	.396**	.252**	.525**	.193**	.577**	.699**	.688**	.668**	.704**	.720**	.677**	.693**	.315**	.559**	.685**	.584**	.638**	.624**	.736**	.791**	1			
23 Negative relationships	.580**	.459**	.323**	.558**	.205**	.530**	.616**	.601**	.603**	.605**	.651**	.564**	.591**	.336**	.468**	.685**	.536**	.640**	.618**	.680**	.666**	.722**	1		
24 Self-defeating	.573**	.461**	.409**	.625**	.310**	.547**	.575**	.598**	.622**	.602**	.645**	.598**	.605**	.297**	.495**	.573**	.517**	.545**	.542**	.709**	.710**	.673**	.619**	1	
25 Self-injury	.576**	.465**	.310**	.429**	.205**	.595**	.673**	.679**	.681**	.671**	.679**	.640**	.683**	.312**	.549**	.597**	.526**	.585**	.589**	.693**	.758**	.656**	.577**	.618**	1

*Key: * indicates a correlation significant (two-tailed) at alpha level 0.05; ** indicates a correlation significant (two tailed) at alpha level 0.01*

Correlations between symptoms and childhood trauma

Spearman's rho correlation coefficients of the relationships between the CTQ subscales, total CTQ score and symptoms are shown in Table 4.

Table 4

Correlations between CTQ subscales, CTQ total score and symptoms

Symptom	CTQ physical abuse subscale	CTQ sexual abuse subscale	CTQ emotional abuse subscale	CTQ physical neglect subscale	CTQ emotional neglect subscale	CTQ total
Hostility	.332**	.287**	.457**	.373**	.356**	.455**
Aggression	.305**	.273**	.358**	.307**	.220**	.355**
Antisocial behaviour	.293**	.207**	.255**	.297**	.231**	.312**
Impulsivity	.217**	.194**	.368**	.310**	.302**	.354**
Callous/ unemotional	.179**	.104*	.175**	.241**	.231**	.242**
Somatising	.387**	.392**	.528**	.393**	.373**	.520**
Interpersonal sensitivity	.324**	.295**	.524**	.366**	.391**	.495**
Depression	.357**	.318**	.522**	.385**	.417**	.509**
Anxiety	.364**	.348**	.554**	.420**	.403**	.530**
Specific phobia	.335**	.328**	.503**	.383**	.386**	.492**
Post-traumatic stress	.360**	.388**	.546**	.446**	.431**	.549**
Obsessive compulsivity	.373**	.317**	.547**	.387**	.396**	.519**
Psychoticism	.400**	.382**	.571**	.421**	.434**	.559**
Magic thinking	.223**	.253**	.307**	.221**	.172**	.286**
Unusual perceptions	.343**	.328**	.476**	.339**	.327**	.454**
Suspiciousness	.396**	.320**	.531**	.361**	.409**	.517**
Ideas of reference	.308**	.265**	.462**	.289**	.308**	.417**
Paranoid thoughts	.388**	.294**	.533**	.418**	.396**	.518**
Persecutory thoughts	.421**	.319**	.558**	.414**	.387**	.535**
Affective instability	.258**	.259**	.478**	.355**	.361**	.439**
Emotional dysregulation	.268**	.279**	.529**	.355**	.428**	.484**
Identity problems	.232**	.286**	.476**	.291**	.327**	.423**
Negative relationships	.275**	.290**	.480**	.305**	.329**	.433**
Self-defeating	.214**	.260**	.389**	.303**	.304**	.376**
Self-injury	.371**	.370**	.559**	.369**	.411**	.526**

Key: * indicates a correlation significant (two-tailed) at alpha level 0.05; ** indicates a correlation significant (two tailed) at alpha level 0.01

Correlation between symptoms and reflective function

Spearman's rho correlation coefficients of the relationship between RFQ_C and RFQ_U and each of the symptoms are shown in Table 5.

Table 5

Correlations between symptoms and RFQ subscales

Symptom	RFQ certainty subscale	RFQ uncertainty subscale
Hostility	-.405**	.548**
Aggression	-.295**	.445**
Antisocial behaviour	-.208**	.308**
Impulsivity	-.399**	.441**
Callous/ unemotional	-.251**	.251**
Somatising	-.330**	.435**
Interpersonal sensitivity	-.469**	.595**
Depression	-.459**	.570**
Anxiety	-.436**	.558**
Specific phobia	-.472**	.587**
Post-traumatic stress	-.396**	.564**
Obsessive compulsivity	-.453**	.547**
Psychoticism	-.453**	.566**
Magic thinking	-.099*	.222**
Unusual perceptions	-.289**	.415**
Suspiciousness	-.441**	.549**
Ideas of reference	-.313**	.403**
Paranoid thoughts	-.390**	.533**
Persecutory thoughts	-.417**	.538**
Affective instability	-.461**	.617**
Emotional dysregulation	-.569**	.691**
Identity problems	-.518**	.586**
Negative relationships	-.426**	.519**
Self-defeating	-.405**	.514**
Self-injury	-.460**	.576**

Key: * indicates a correlation significant (two-tailed) at alpha level 0.05;

** indicates a correlation significant (two tailed) at alpha level 0.01

Models of symptom comorbidity

Model fit

The model fit statistics for each of the models tested are shown in Table 6. The full bifactor model was the best fit for the data and approached an adequate fit. It was also the preferred model according to the parsimony indicators of AIC and ABIC.

Factor loadings

Standardised factor loadings are shown for the group factor, bifactor and modified bifactor with no internalising group factor (Model 6) models in Table 8. Although the group factor model was not a good fit for the data overall (Table 6), the factor loadings were of a reasonable magnitude, providing support for the group factor structure chosen. Although the full bifactor model was the best fitting model overall, parameter estimates for several INT symptoms were not significant and so the Model 6 factor loadings are included for comparison.

Relationships between factors

The correlations between group factors for the full bifactor and the group factor models are shown in Table 7, demonstrating the effect of controlling for *p* on inter-factor correlations.

Table 6

Fit and parsimony indices for alternative models of the comorbidity structure of symptomatology

Model and number of parameters		Chi-sq	df	CFI	TLI	RMSEA [90% CI]	SRMR	AIC	ABIC[+]
Model 1: Unidimensional	75	2252.39	275	0.805	0.787	0.119 [0.115, 0.124]	0.067	67004.4	67083.2
Model 2: Group factor	81	1505.17	269	0.878	0.864	0.095 [0.091, 0.100]	0.058	66155.2	66240.3
Model 3: Full bifactor	**106**	**1105.68**	**244**	**0.915**	**0.895**	**0.084 [0.079, 0.089]**	**0.039**	**65686.8**	**65798.1**
Model 4: Modified bifactor (no TD)	96	1440.59	254	0.883	0.862	0.096 [0.091, 0.101]	0.047	66033.9	66134.7
Model 5: Modified bifactor (no BOR)	97	1246.11	253	0.902	0.884	0.088 [0.083, 0.093]	0.045	65986.5	65884.6
Model 6: Modified bifactor (no INT)	96	1177.16	254	0.909	0.892	0.085 [0.080, 0.090]	0.041	65803.6	65904.5

Abbreviations: ABIC: Bayesian Information Criterion (sample-size adjusted); AIC: Akaike's Information Criterion; CFI = comparative fit index; df = degrees of freedom; SRMR = Standardized Root Mean Square Residual; RMSEA = root mean square error of approximation; TLI = Tucker–Lewis index

Table 7

Inter-factor correlations for group and bifactor models

	Group factor model (Model 2)				Bifactor model (Model 3)			
	ASOC	**INT**	**TD**	**BOR**	**ASOC**	**INT**	**TD**	**BOR**
ASOC	1				1			
INT	0.72**	1			0.37 (ns)	1		
TD	0.81**	0.95**	1		0.54**	0.64**	1	
BOR	0.76**	0.87**	0.84**	1	0.51**	0.46 (ns)	0.46**	1

Key: * indicates a correlation significant (two-tailed) at alpha level 0.05

** indicates a correlation significant (two tailed) at alpha level 0.01

Table 8

Symptom standardised factor loadings across group factor and selected bifactor models

Symptom	Group factor model		Bifactor model (Model 3)			Modified bifactor model (Model 6)		
	Factor loading	Residual variance	Factor loading		Residual variance	Factor loading		Residual variance
	On ASOC		*On ASOC*	*On p*		*On ASOC*	*On p*	
Hostility	0.824**	0.322**	0.422**	0.696**	0.337**	0.355**	0.724**	0.349**
Aggression	0.722**	0.479**	0.736**	0.369**	0.322**	0.694**	0.438**	0.326**
Antisocial behaviour	0.585**	0.658**	0.668**	0.241**	0.496**	0.650**	0.301**	0.487**
Impulsivity	0.736**	0.458**	0.528**	0.488**	0.483**	0.490**	0.526**	0.483**
Callous/ unemotional	0.456**	0.792**	0.381**	0.241**	0.797**	0.380**	0.255**	0.791**
	On INT		*On INT*					
Somatising	0.876**	0.373**	0.285**	0.706**	0.329**		0.783**	0.387**
Interpersonal sensitivity	0.792**	0.204**	0.415**	0.895**	0.183**		0.897**	0.195**
Depression	0.892**	0.211**	0.122 (ns)	0.924**	0.144**		0.901**	0.188**
Anxiety	0.888**	0.141**	0.048 (ns)	0.868**	0.125**		0.923**	0.148**
Specific phobia	0.927**	0.240**	0.349**	0.810**	0.233**		0.866**	0.250**
Post-traumatic stress	0.872**	0.291**	0.334**	0.775**	0.290**		0.834**	0.305**
Obsessive compulsivity	0.842**	0.232**	0.331 (ns)	0.829**	0.232**		0.872**	0.239**
	On TD		*On TD*			*On TD*		
Psychoticism	0.884**	0.219**	0.121 (ns)	0.919**	0.141**	0.007 (ns)	0.911**	0.170**
Magic thinking	0.483**	0.767**	0.618**	0.274**	0.543**	0.533**	0.372**	0.577**
Unusual perceptions	0.698**	0.513**	0.599**	0.528**	0.362**	0.479**	0.622**	0.384**
Suspiciousness	0.816**	0.335**	0.445**	0.703**	0.308**	0.370**	0.753**	0.297**
Ideas of reference	0.744**	0.447**	0.584**	0.584**	0.319**	0.527**	0.660**	0.286**
Paranoid thoughts	0.895**	0.199**	0.304**	0.820**	0.235**	0.176**	0.851**	0.244**
Persecutory thoughts	0.846**	0.284**	0.387**	0.736**	0.308**	0.264**	0.782**	0.318**

Symptom	Group factor model		Bifactor model (Model 3)			Modified bifactor model (Model 6)		
	Factor loading	Residual variance	Factor loading		Residual variance	Factor loading		Residual variance
	On BOR		*On BOR*			*On BOR*	*On p*	
Affective instability	0.871**	0.242**	0.554**	0.688**	0.220**	0.487**	0.735**	0.224**
Emotional dysregulation	0.919**	0.155**	0.441**	0.797**	0.170**	0.391**	0.821**	0.174**
Identity problems	0.869**	0.246**	0.493**	0.716**	0.244**	0.452**	0.747**	0.238**
Negative relationships	0.779**	0.393**	0.476**	0.629**	0.379**	0.423**	0.667**	0.376**
Self-defeating	0.778**	0.394**	0.503**	0.608**	0.377**	0.439**	0.652**	0.382**
Self-injury	0.805**	0.351**	0.392**	0.698**	0.360**	0.329**	0.727**	0.363**

Key: ** indicates a correlation significant (two tailed) at alpha level 0.01

Psychopathology, childhood trauma and reflective function

Association between latent factors and childhood trauma

Spearman's rho correlation coefficients of the relationship between total CTQ score and factor scores for the group and bifactor model latent factors are shown in Table 9.

Table 9

Association between model latent factors and CTQ total score

	Group factor model (Model 2)		Bifactor model (Model 3)	
	r_s	*p*	r_s	*p*
p	-	-	.549	< .001
ASOC	.533	< .001	.245	< .001
INT	.578	< .001	.276	< .001
TD	.590	< .001	.278	< .001
BOR	.538	< .001	.208	< .001

Association between latent factors and reflective function

Spearman's rho correlation coefficients for the relationship between RFQ_U and RFQ_C, and factor scores for the group and bifactor model latent factors are shown in Table 10.

Table 10

Association between latent factors and sub-scales of RFQ

	RFQ certainty subscale				RFQ uncertainty subscale			
	Group factor (Model 2)		Bifactor (Model 3)		Group factor (Model 2)		Bifactor (Model 3)	
	r_s	*p*	r_s	*p*	r_s	*p*	r_s	*p*
p	-	-	-.500	< .001	-	-	.622	< .001
ASOC	-.467	< .001	-.178	< .001	.617	< .001	.283	< .001
INT	-.494	< .001	-.084	.068	.629	< .001	.191	< .001
TD	-.470	< .001	-.071	.127	.618	< .001	.166	< .001
BOR	-.552	< .001	-.338	< .001	.693	< .001	.367	< .001

Discussion

The comorbidity structure of symptomatology

This study supports previous work identifying a tendency towards experiencing any and all symptoms comorbidly, challenging the supposition that mental health disorders are comorbid at random. The full bifactor model approached an adequate fit for the data (Table 6), which is particularly remarkable in view of the fact that a broader range of symptoms were modelled here than by previous studies. The unifactorial model was a poor fit for the data and the bifactor models (Model 3, Model 5 and Model 6, and, according to some indices, Model 4) were a better fit for the data than the correlated group factor model, indicating that both general and specific sources of variation influenced the pattern of comorbidity observed. Of the modified bifactor models, Model 6 (with no INT group factor) fitted the data better than the models where the more severe thought disorder and borderline symptoms loading only onto *p* (Models 4 and 5), and possible reasons for this are discussed below.

This study replicated a pattern observed in several other studies whereby correlations between group factors were attenuated when *p* was controlled for in a bifactor model (Table 7; Carragher et al., 2016; Caspi et al., 2014; Laceulle et al., 2015). The larger inter-factor correlations observed in this study than others are likely due to the inflated correlations between symptoms, due to the purposive sampling. Finally, although the robust extraction of dimensional factors across this and other studies supports arguments against sampling from narrow populations (Insel et al., 2010), the sampling technique employed here limits the generalisability of these results.

Characteristics of the p-statistic

In advance of a more substantive discussion of the results it may be helpful to make a distinction between ‘the *p*-factor’, a broad finding across various studies, and a particular ‘*p*-statistic’, which reflects study-specific differences. Although differences in how

psychological constructs are operationalised are ubiquitous in research, this distinction is particularly important in this context, as differences between studies are constitutive of the construct (*p*) measured. This study is one of several that share the same broad aims and method, and from a high-level perspective the extraction of 'the *p*-factor' across such studies is striking; however, differences between '*p*-statistics' preclude strong conclusions being drawn. Such differences between studies include, but are not limited to, symptoms measured, population characteristics, model characteristics (for example, factor specification), general characteristics of the CFA modelling technique (for example, its simplifying assumptions) and interactions between these factors. A discussion of trends observed must therefore be tentative.

In the full bifactor model, symptoms which appeared to be indicators of *p* due to high loadings are interpersonal sensitivity, depression, anxiety and psychoticism, and, to a lesser extent, phobia, obsessive-compulsivity, post-traumatic stress, paranoid thoughts and emotional dysregulation (Table 8). Broadly, these symptoms could be considered indicative of heighted sensitivity, distress and anxious thought disturbance. These results should be interpreted with awareness that restricted cross-loadings in CFA inflate loadings on a general factor (Murray & Johnson, 2013), which may be of particular relevance here as both borderline and internalising symptomatology, which have overlapping diagnostic criteria, were modelled in a sample with a high proportion of people with borderline presentations. Nonetheless, these results seem to accord with a previous finding in an adult sample that depression, generalised anxiety, phobia, obsessive-compulsivity, mania and schizophrenia all load more heavily on *p* than their respective group factors (Caspi et al., 2014). Studies investigating *p* in child and adolescent samples also found results which accord with those here, with distress-type symptoms of unhappiness, being nervous in new situations (Carragher et al., 2016) and depression (Martel et al., 2017) loading highly on p.

Symptoms which had low loadings on *p* (smaller than .4) were aggression and antisocial behaviour, and callous-unemotional traits and magical thinking, which also had low

loadings on their respective group factors. This pattern may be partly due to characteristics of these symptoms; although there is limited evidence from adult samples (Caspi and colleagues' externalising factor was largely specified by substance misuse), aggression and delinquency loaded less highly on *p* than on an externalising factor in several studies with child and adolescent samples (Carragher et al., 2016; Laceulle et al., 2015). The pattern observed may also be partly due to characteristics of the symptoms as they were measured here; for example, the callous/unemotional measure includes items about personal charm and concern for others, and the magical thinking measure asks about clairvoyance, telepathy and astrology (Appendix B6), which might be associated with feelings of personal efficacy rather than distress. However, these findings should be interpreted with particular caution due to a limitation of this study, that callous/unemotional traits and impulsivity were measured using a scale which has not been validated in adult populations and the internal consistency of the callous/emotional subscale was poor (Appendix B6).

With the caveat that the group factor model was not a good fit for the data (Table 6), there were several trends observed when factor loadings for this model were compared with the full bifactor model (Table 8). Several symptoms continued to load more heavily on their group factor than p, indicating they tended to co-occur with symptoms of a similar type; at the group level, ASOC symptoms tended to follow this pattern. In addition to the 'high-p' symptoms discussed above, symptoms which loaded more heavily on p, and in some cases only marginally on their group factor when *p* was controlled for, included the remaining INT symptoms, hostility, self-injury, persecutory thoughts and suspiciousness. The borderline symptoms also loaded highly on p, but retained robust loadings on BOR. It is difficult to interpret these results as they might be affected by the high prevalence of borderline personality disorder in the sample, however, they appear to show that symptoms on a broad borderline spectrum tended to be comorbid with a range of other symptoms. The TD symptoms tended to be split between those which continued to load more heavily on TD

(unusual perceptions and magical thinking) and the more paranoid-type thought disorder symptoms, which loaded more heavily on *p*.

Psychopathology, childhood trauma and reflective function

Internalised representations of childhood maltreatment were significantly associated with every symptom measured (Table 4) as were both certainty and uncertainty about mental states (Table 5). These findings cohere with the latent vulnerability theory of childhood maltreatment and mentalising theory, which suggest that impairments in certain psychological processes may leave individuals vulnerable to stressors and a range of mental health problems (Fonagy et al., 2016; McCrory & Viding, 2015). These findings also support the analytic strategy of this study, of investigating the relationships between these variables and the higher-order factor, *p*.

The correlations between childhood maltreatment and latent factor scores (Table 9) replicated the pattern observed in a previous study (Caspi et al., 2014), whereby childhood maltreatment was correlated highly with *p* and its correlations with group factors were attenuated when *p* was controlled for. This study is the first to examine the relationship between reflective function and *p*. Both certainty and uncertainty about mental states also correlated highly with *p*, and their correlations with group factors were attenuated when *p* was controlled for. However, as would be expected in light of mentalising theory, the correlations with BOR were larger than with other group factors, even when *p* was controlled for (Table 10). These results at the factor-level reflect the strong associations (approaching 0.5 or above) between childhood maltreatment and individual 'high-p' symptoms (Table 4), which fits with evidence that maltreatment is associated with mood disorders (McCrory & Viding, 2015). Similarly, the strongest associations between impairments in reflective function were with borderline symptoms and 'high-*p*' symptoms (Table 5), fitting with accounts linking impairments in mentalising with interpersonal-sensitivity, paranoid anxiety, distress (Fonagy, 1999) and depression (Fonagy et al., 2016).

This study is important corroboration of the lack of specificity between psychopathology and both child maltreatment and reflective function, as well as evidence of their association with a tendency towards comorbid presentations. The results demonstrate that useful information can be gained through examination of risk factors for psychopathology at different levels of generality. However, there are several study limitations which temper what can be concluded on the basis of these results. Firstly, the CTQ may overestimate childhood trauma (McCrory, Gerin, & Viding, 2017). Secondly, as a retrospective rather than actuarial measure, the CTQ measures what people have internalised about past; and with regards both internalised representations of child trauma and mentalising ability, causation cannot be established and it is possible that psychopathology influenced responses on these measures. However, these results suggest that a priority for future research could be differentiating specific mechanisms conferring risk for psychopathology (Cecil, Viding, Fearon, Glaser, & McCrory, 2017) and investigating path effects within different modelling frameworks; for example, investigating whether mentalising ability partially mediates the relationship between childhood trauma and *p* (Fonagy, Gergely, Jurist, & Target, 2004).

The p-factor in context

One of the findings of this study was that the fit of the modified bifactor model, with no INT group factor, approximated that of the full bifactor model (Table 6). That several of the internalising symptoms (with psychoticism) did not load significantly onto INT in the full bifactor model (Table 8) implies that, statistically, the modified bifactor could be a contender for best-describing the data. This raises an interesting question, of what the qualitative status of 'high-p' symptoms is. Several of the seminal papers in this area found that symptoms which are qualitatively severe also loaded highly on *p* (Caspi et al., 2014; Sharp et al., 2015); extrapolating from this, others have considered that high-p symptoms, by virtue of the fact they load highly on p, are 'more pathological' (Laceulle et al., 2015). Caspi and colleagues (2014) questioned whether the *p*-factor might be a statistical *reductio ad absurdum*; and a

relevant question for this study, in which both common and severe symptoms were modelled, might be whether there results constitute a *reductio ad absurdum* for a naïve interpretation of high-*p* symptoms as especially 'pathological'.

To interpret these findings in a clinically-meaningful way it is important to place them in context. In some respects the empirical approach to nosology departs from the diagnostic system, however, it also retains certain assumptions. Within the diagnostic system, diagnoses are categorical; however, there is a *de facto* ranking in terms of severity. Proponents of the empirical approach argue that important information is lost through the dichotomous measurement of dimensional symptoms (Kotov et al., 2017), however, information regarding the ranking of diagnoses or symptoms, in terms of severity, is also lost when they are placed on the same rubric (a covariance matrix). To illustrate this point; self-report scales, such as those used in this study, are intended to measure a particular symptom avoiding 'floor' or 'ceiling' effects (Bech, 2012). High scores on scales measuring a simple phobia and schizophrenia might measure severity *within each category* in a meaningfully way, but they do not both measure the same degree of impairment. This implies the covariance matrices on which CFA is based cannot be interpreted as straightforwardly representing severity or 'pathology'.

Interestingly, this observation may shed light on the fact that the two studies which found thought disorder did not form a separate group factor used longitudinal data (Caspi et al., 2014; Laceulle et al., 2015), as thought disorders may be enduring and sequentially comorbid with anxiety and depression during the prodromal period (Meyer et al., 2005). Similarly, personality disorders are persistent presentations, and both may look different, for example, to episodic depression. The introduction of the additional dimension of time may therefore mitigate against the fact that cross-sectional variability is not fully described by CFA models. Therefore these results, which are derived from cross-sectional data, are not evidence against the structural hypothesis. (Although Sharp and colleagues (2015) did not use longitudinal data, as only personality disorders were modelled it could be speculated

that there was less variation both between cross-sectionally measured severity and typical duration of presentation.) How best to think about this psychometric artefact of the diagnostic system is a general question for an empirical nosology, however, it is a particular issue for this cross-sectional study of common and severe presentations. Statistical modelling involves simplification, which will inevitably mean that models are 'wrong' to some degree (Eaton, 2015), and additional clinically-informed consideration may be required for interpreting such limitations.

Clinical implications

Conceptualising p at the level of the individual

An important question for empirical nosological research is how best to interpret transdiagnostic population statistics. A criticism levied at the diagnostic system is that people with the same diagnosis may have different symptoms (Beutler & Malik, 2002); however, it might be that the only way to operationalise factors at a higher level of abstraction than symptoms (diagnoses and transdiagnostic factors) at the level of the individual is polythetically. For example, there might be more than 10,000 ways for a person to have PTSD according to DSM-5 criteria (Rosen, Lilienfeld, Frueh, McHugh, & Spitzer, 2010), but there will be even more ways for them to be 'high-INT', and yet again more ways to be 'high-*p*'.

Transdiagnostic treatments

The authors of several studies have suggested that the *p*-factor might support the unified protocols approach (Caspi et al., 2014; Subica et al., 2015), which adapts cognitive behavioural therapy (CBT) for presentations that do not fit diagnostic categories (Barlow, Allen, & Choate, 2004). Symptom-focussed approaches such as these face the question of which symptoms to target first (Butler, Fennell, & Hackmann, 2010), and future research might fruitfully investigate whether targeting 'high-p' symptoms improves outcomes.

Taken in conjunction with mentalising theory and other evidence, the results of this study suggest an alternative formulation of transdiagnostic treatment, as therapy aimed at improving mentalising capacity. Such treatments might include Mentalization Based Treatment (MBT), which has been adapted for a range of presentations (Bateman & Fonagy, 2012). However, it has also been suggested that a 'common mechanism' across a range of evidence-based therapies is the stimulation of mentalising ability (Fonagy, Luyten, & Allison, In Press). The treatment implications of this research require further investigation, but transdiagnostic factors may offer a framework within which to formulate and test therapeutic mechanisms of change.

Finally, the results here are evidence in support of a stress-diathesis model whereby general psychological vulnerabilities interact with other stressors to influence specific symptoms. Proponents of the latent vulnerability model of childhood maltreatment suggest early intervention in cases where victims of maltreatment are not (yet) symptomatic (McCrory & Viding, 2015). Extrapolating from this suggestion, these findings might lend broad support for early intervention models for those at risk of psychopathology.

Transdiagnostic assessment

One implication of the first conception, outlined above, of the 'transdiagnostic' in terms of variability in symptom patterns, could be to operationalise transdiagnostic factors in terms of polythetic symptom criteria. This might imply making comorbidity a more specific focus of broad assessments of mental health. Alternatively, 'high-p' symptoms (Patalay et al., 2015) or risk factors could be used as indicators of potentially raised risk of comorbid psychopathology, which, in practice, might be translated into algorithmic clinical decision-making tools. On the second conception of transdiagnostic factors, as reflecting broad psychological processes, assessment might more usefully focus on these factors. Taken in conjunction with mentalising theory, these results indicate that an assessments of mentalising ability, already part of MBT (Bateman & Fonagy, 2012), might be usefully incorporated into a wider range of assessment contexts.

Evaluation of the study

A strength of this study is that, to the author's knowledge, it is the first to investigate the structure of both Axis I and borderline symptomatology in a bifactor model. Although such models may require careful interpretation, that such a broad '*p*-statistic' was extracted is particularly remarkable. Another strength of the study is that it examined the statistical relationships between particular symptoms and latent factors which describe them. Although these relationships may reflect spurious influences, a better understanding of potential sources of variation will be gained through comparisons across studies. Finally, bifactor modelling allows for the investigation of risk factors which have general effects, offering an opportunity for nosological research to investigate theories of how symptoms might arise in the mind.

This study also had a number of limitations. The purposive sampling technique limits the generalisability of the results and the exclusion of people with active psychosis may have resulted in other thought disorder symptoms being underrepresented. The overlap between the clinical and control groups on borderline symptoms measured by the PAI-BOR was unexpected, and might indicate either the unreliability of this self-report measure or clinical levels of personality pathology in some of the control participants, or potentially both. Although self-report scales are limited generally (Stone et al., 1999), there may be particular constructs which are less amenable to this method of data collection. For example, the challenge of developing a self-report scale measuring mentalising is explored by the developers of the RFQ (Fonagy et al., 2016), the CTQ may overestimate child trauma (McCrory et al., 2017) and it the measurement of certain symptoms, such as thought disorder, may be particularly susceptible to bias. There were also specific limitations; the APSD is not validated in adult samples and one of its subscales had poor internal consistency, and a number of the measures used were brief screening tools (Appendix B6). Finally, although another study has examined the criterion validity of the *p*-factor in

adolescence by comparing self-report data with teacher report (Lahey et al., 2015), it is possible that the *p*-statistic measured here represents a scoring bias.

There were also several limitations related to the study design and analysis. Firstly, cross-sectional symptom-based data is a narrow way of conceptualising psychopathology and covariance matrices do not fully reflect clinically-meaningful differences between symptomatic presentations, which might be better-captured using a longitudinal design. This is a broad issue facing empirical nosological research, as the majority of studies are cross-sectional, and addressing this should be a priority. Secondly, the cross-sectional design meant that causal effects of childhood trauma and mentalising on psychopathology could not be established, a question which could be addressed by future research within a prospective design. Thirdly, all modelling involves methodological decisions which potentially bias the results. Decisions regarding the specification of group factors were made, where possible, on the basis of prior research. However, given the range of symptoms modelled there is a strong argument for prior exploratory analysis of this set of symptoms (Reise, 2012) or an analytical approach such as exploratory structural equation modelling which would allow aspects of the model based on robust findings to be fixed whilst others are free (Wright, 2017). These limitations highlight that the empirical nosological approach is imperfect, its findings require careful clinical consideration and it cannot sidestep difficult qualitative questions; however, it may offer new and generative insights into psychopathology which move nosology beyond the limitations of the categorical diagnostic system.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.

Association, A. P., Statistics, A. P. A. T. F. on N. and, & Spitzer, R. L. (1980). *Diagnostic and Statistical Manual of Mental Disorders: DSM III.* American Psychiatric Association.

Barlow, D. H., Allen, L. B., & Choate, M. L. (2004). Toward a unified treatment for emotional disorders. *Behavior Therapy*, *35*(2), 205–230.

Bateman, A., & Fonagy, P. (2012). *Handbook of Mentalizing in Mental Health Practice*. American Psychiatric Pub.

Bech, P. (2012). *Clinical Psychometrics*. John Wiley & Sons.

Bentall, R. P. (2004). *Madness explained: Psychosis and human nature*. Penguin UK.

Bernstein, D. P., Fink, L., Handelsman, L., Foote, J., Lovejoy, M., Wenzel, K., … Ruggiero, J. (1994). Initial reliability and validity of a new retrospective measure of child abuse and neglect. *The American Journal of Psychiatry*, *151*(8), 1132.

Beutler, L. E., & Malik, M. L. (2002). *Rethinking the DSM: A Psychological Perspective*. American Psychological Association.

Brodbeck, J., Stulz, N., Itten, S., Regli, D., Znoj, H., & Caspar, F. (2014). The structure of psychopathological symptoms and the associations with DSM-diagnoses in treatment seeking individuals. *Comprehensive Psychiatry*, *55*(3), 714–726.

Butler, G., Fennell, M., & Hackmann, A. (2010). *Cognitive-Behavioral Therapy for Anxiety Disorders: Mastering Clinical Challenges*. Guilford Press.

Carragher, N., Teesson, M., Sunderland, M., Newton, N. C., Krueger, R. F., Conrod, P. J., … Slade, T. (2016). The structure of adolescent psychopathology: a symptom-level analysis. *Psychological Medicine*, *46*(05), 981–994.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., … Poulton, R. (2014). The p factor one general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, *2*(2), 119–137.

Cecil, C. A., Viding, E., Fearon, P., Glaser, D., & McCrory, E. J. (2017). Disentangling the mental health impact of childhood abuse and neglect. *Child Abuse & Neglect*, *63*, 106–119.

Coccaro, E. F., Berman, M. E., & Kavoussi, R. J. (1997). Assessment of life history of aggression: development and psychometric characteristics. *Psychiatry Research*, *73*(3), 147–157.

Debbané, M., Salaminios, G., Luyten, P., Badoud, D., Armando, M., Tozzi, A. S., … Brent, B. K. (2016). Attachment, neurobiology, and mentalizing along the psychosis continuum. *Frontiers in Human Neuroscience*, *10*.

Derogatis, L. R., & Melisaratos, N. (1983). The brief symptom inventory: an introductory report. *Psychological Medicine*, *13*(03), 595–605.

Eaton, N. R. (2015). Latent variable and network models of comorbidity: toward an empirically derived nosology. *Social Psychiatry and Psychiatric Epidemiology*, *50*(6), 845–849.

Enders, C. K. (2003). Using the Expectation Maximization Algorithm to Estimate Coefficient Alpha for Scales With Item-Level Missing Data. *Psychological Methods*, *8*(3), 322–337. https://doi.org/10.1037/1082-989X.8.3.322

Fonagy, P. (1991). Thinking about Thinking: Some Clinical and Theoretical Considerations in the Treatment of a Borderline Patient. *Int. J. Psycho-Anal.*, *72*, 639–656.

Fonagy, P. (1999). Attachment, the development of the self, and its pathology in personality disorders. In *Treatment of personality disorders* (pp. 53–68). Springer.

Fonagy, P., Gergely, G., Jurist, E. L., & Target, M. (2004). *Affect Regulation, Mentalization and the Development of the Self*. Karnac Books.

Fonagy, P., Luyten, P., & Allison, E. (In Press). Epistemic petrification and the restoration of epistemic trust: A new conceptualization of borderline personality disorder and its psychosocial treatment. *Journal of Personality Disorders*.

Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y.-W., Warren, F., Howard, S., … Lowyck, B. (2016). Development and Validation of a Self-Report Measure of Mentalizing: The Reflective Functioning Questionnaire. *PLOS ONE*, *11*(7), e0158678. https://doi.org/10.1371/journal.pone.0158678

Geiser, C. (2013). *Data Analysis with Mplus* (1st edition). New York: Guilford Press.

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in

emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment*, *26*(1), 41–54.

Green, C. E. L., Freeman, D., Kuipers, E., Bebbington, P., Fowler, D., Dunn, G., & Garety, P. A. (2008). Measuring ideas of persecution and social reference: the Green et al. Paranoid Thought Scales (GPTS). *Psychological Medicine*, *38*(01), 101–111.

Grover, K. E., Carpenter, L. L., Price, L. H., Gagne, G. G., Mello, A. F., Mello, M. F., & Tyrka, A. R. (2007). The Relationship Between Childhood Abuse and Adult Personality Disorder Symptoms. *Journal of Personality Disorders*, *21*(4), 442–447. https://doi.org/10.1521/pedi.2007.21.4.442

Hoertel, N., Franco, S., Wall, M. M., Oquendo, M. A., Kerridge, B. T., Limosin, F., & Blanco, C. (2015). Mental disorders and risk of suicide attempt: a national prospective study. *Molecular Psychiatry*, *20*(6), 718–726.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., … Wang, P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *American Journal of Psychiatry*, *167*(7), 748–751.

Jahng, S., Trull, T. J., Wood, P. K., Tragesser, S. L., Tomko, R., Grant, J. D., … Sher, K. J. (2011). Distinguishing general and specific personality disorder features and implications for substance dependence comorbidity. *Journal of Abnormal Psychology*, *120*(3), 656.

Kim, H., & Eaton, N. R. (2015). The hierarchical structure of common mental disorders: Connecting multiple levels of comorbidity, bifactor models, and predictive validity. *Journal of Abnormal Psychology*, *124*(4), 1064.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., … Clark, L. A. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A

dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454.

Kotov, R., Ruggero, C. J., Krueger, R. F., Watson, D., Yuan, Q., & Zimmerman, M. (2011). New Dimensions in the Quantitative Classification of Mental Illness. *Archives of General Psychiatry*, *68*(10), 1003–1011. https://doi.org/10.1001/archgenpsychiatry.2011.107

Krueger, R. F. (1999). The structure of common mental disorders. *Arch Gen Psychiat*, *56*, 921–926.

Krueger, R. F., & Markon, K. E. (2011). A dimensional-spectrum model of psychopathology: progress and opportunities. *Archives of General Psychiatry*, *68*(1), 10–11.

Krueger, R. F., & Markon, K. E. (2014). The role of the DSM-5 personality trait model in moving toward a quantitative and empirically based approach to classifying personality and psychopathology. *Annual Review of Clinical Psychology*, *10*, 477–501. https://doi.org/10.1146/annurev-clinpsy-032813-153732

Laceulle, O. M., Vollebergh, W. A., & Ormel, J. (2015). The Structure of Psychopathology in Adolescence Replication of a General Psychopathology Factor in the TRAILS Study. *Clinical Psychological Science*, 2167702614560750.

Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is There a General Factor of Prevalent Psychopathology during Adulthood? *Journal of Abnormal Psychology*, *121*(4), 971–977. https://doi.org/10.1037/a0028355

Lahey, B. B., Rathouz, P. J., Keenan, K., Stepp, S. D., Loeber, R., & Hipwell, A. E. (2015). Criterion validity of the general factor of psychopathology in a prospective study of girls. *Journal of Child Psychology and Psychiatry*, *56*(4), 415–422.

Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data*. John Wiley & Sons.

Luyten, P., Van Houdenhove, B., Lemma, A., Target, M., & Fonagy, P. (2012). A mentalization-based approach to the understanding and treatment of functional somatic disorders. *Psychoanalytic Psychotherapy*, *26*(2), 121–140.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130.

Martel, M. M., Pan, P. M., Hoffmann, M. S., Gadelha, A., do Rosário, M. C., Mari, J. J., … Bressan, R. A. (2017). A general psychopathology factor (P factor) in children: Structural model analysis and external validation through familial risk and child global executive function. *Journal of Abnormal Psychology*, *126*(1), 137.

McCrory, E. J., Gerin, M. I., & Viding, E. (2017). Annual Research Review: Childhood maltreatment, latent vulnerability and the shift to preventative psychiatry–the contribution of functional brain imaging. *Journal of Child Psychology and Psychiatry*, *58*(4), 338–357.

McCrory, E. J., & Viding, E. (2015). The theory of latent vulnerability: Reconceptualizing the link between childhood maltreatment and psychiatric disorder. *Development and Psychopathology*, *27*(02), 493–505.

McDonald, S. D., & Calhoun, P. S. (2010). The diagnostic accuracy of the PTSD checklist: a critical review. *Clinical Psychology Review*, *30*(8), 976–987.

Meyer, S. E., Bearden, C. E., Lux, S. R., Gordon, J. L., Johnson, J. K., O'Brien, M. P., … Cannon, T. D. (2005). The psychosis prodrome in adolescent patients viewed through the lens of DSM-IV. *Journal of Child & Adolescent Psychopharmacology*, *15*(3), 434–451.

Miller, M. W., Wolf, E. J., Reardon, A., Greene, A., Ofrat, S., & McInerney, S. (2012). Personality and the latent structure of PTSD comorbidity. *Journal of Anxiety Disorders*, *26*(5), 599–607. https://doi.org/10.1016/j.janxdis.2012.02.016

Moray, L. (1991). *Personality Assessment Inventory Professional Manual*. Psychological Assessment Resources, Inc.

Munoz, L. C., & Frick, P. J. (2007). The reliability, stability, and predictive utility of the self-report version of the antisocial process screening device. *Scandinavian Journal of Psychology*, *48*(4), 299–312.

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence*, *41*(5), 407–422. https://doi.org/10.1016/j.intell.2013.06.004

Muthén, L. K., & Muthén, B. O. (2006). Mplus Version 7 user's guide. *Los Angeles, CA: Muthén & Muthén*.

Nanni, V., Uher, R., & Danese, A. (2012). Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: a meta-analysis. *American Journal of Psychiatry*, *169*(2), 141–151.

Nolte, T., Guiney, J., Fonagy, P., Mayes, L. C., & Luyten, P. (2011). Interpersonal stress regulation and the development of anxiety disorders: an attachment-based developmental framework. *Frontiers in Behavioral Neuroscience*, *5*, 55.

O'Connor, B. P. (2005). A search for consensus on the dimensional structure of personality disorders. *Journal of Clinical Psychology*, *61*(3), 323–345.

Patalay, P., Fonagy, P., Deighton, J., Belsky, J., Vostanis, P., & Wolpert, M. (2015). A general psychopathology factor in early adolescence. *The British Journal of Psychiatry*, *207*(1), 15. https://doi.org/10.1192/bjp.bp.114.149591

Raine, A. (1991). The SPQ: a scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophrenia Bulletin*, *17*(4), 555.

Raine, A., Reynolds, C., Lencz, T., Scerbo, A., Triphon, N., & Kim, D. (1994). Cognitive-perceptual, interpersonal, and disorganized features of schizotypal personality. *Schizophrenia Bulletin*, *20*(1), 191.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696.

Released, I. C. (2016). IBM SPSS Statistics for Windows. 24. *Armonk, NY: IBM Corp*.

Rosen, G. M., Lilienfeld, S. O., Frueh, B. C., McHugh, P. R., & Spitzer, R. L. (2010). Reflections on PTSD's future in DSM–V. *The British Journal of Psychiatry*, *197*(5), 343–344.

Salsman, N. L., & Linehan, M. M. (2012). An investigation of the relationships among negative affect, difficulties in emotion regulation, and features of borderline personality disorder. *Journal of Psychopathology and Behavioral Assessment*, *34*(2), 260–267.

SAS Institute. (n.d.). *SAS*.

Scott, K. M., Smith, D. R., & Ellis, P. M. (2010). Prospectively ascertained child maltreatment and its association with DSM-IV mental disorders in young adults. *Archives of General Psychiatry*, *67*(7), 712–719.

Sharp, C., Wright, A. G., Fowler, J. C., Frueh, B. C., Allen, J. G., Oldham, J., & Clark, L. A. (2015). The structure of personality pathology: Both general (‘g’) and specific (‘s’) factors? *Journal of Abnormal Psychology*, *124*(2), 387.

Skodol, A. E., Gunderson, J. G., Pfohl, B., Widiger, T. A., Livesley, W. J., & Siever, L. J. (2002). The borderline diagnosis I: psychopathology, comorbidity, and personaltity structure. *Biological Psychiatry*, *51*(12), 936–950.

Stein, M. B., Pinsker-Aspen, J. H., & Hilsenroth, M. J. (2007). Borderline pathology and the Personality Assessment Inventory (PAI): An evaluation of criterion and concurrent validity. *Journal of Personality Assessment*, *88*(1), 81–89.

Stochl, J., Khandaker, G. M., Lewis, G., Perez, J., Goodyer, I. M., Zammit, S., … Jones, P. B. (2015). Mood, anxiety and psychotic phenomena measure a common psychopathological factor. *Psychological Medicine*, *45*(07), 1483–1493.

Stone, A. A., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (Eds.). (1999). *The Science of Self-report: Implications for Research and Practice* (1 edition). Mahwah, NJ: Psychology Press.

Subica, A. M., Allen, J. G., Frueh, B. C., Elhai, J. D., & Fowler, J. C. (2015). Disentangling depression and anxiety in relation to neuroticism, extraversion, suicide, and self-harm among adult psychiatric inpatients with serious mental illness. *British Journal of Clinical Psychology*.

Varese, F., Smeets, F., Drukker, M., Lieverse, R., Lataster, T., Viechtbauer, W., … Bentall, R. P. (2012). Childhood adversities increase the risk of psychosis: a meta-analysis of patient-control, prospective-and cross-sectional cohort studies. *Schizophrenia Bulletin*, sbs050.

Vollebergh, W. A. M., Iedema, J., Bijl, R. V., Graaf, R. de, Smit, F., & Ormel, J. (2001). The Structure and Stability of Common Mental Disorders: The NEMESIS Study. *Archives of General Psychiatry*, *58*(6), 597–603. https://doi.org/10.1001/archpsyc.58.6.597

Waldman, I. D., Poore, H. E., van Hulle, C., Rathouz, P. J., & Lahey, B. B. (2016). External validity of a hierarchical dimensional model of child and adolescent psychopathology: Tests using confirmatory factor analyses and multivariate behavior genetic analyses. *Journal of Abnormal Psychology*, *125*(8), 1053.

Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993). The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility. In *annual convention of the international society for traumatic stress studies, San Antonio, TX* (Vol. 462). San Antonio, TX.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models an evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, *73*(6), 913–934.

Wolf, E. J., Miller, M. W., & Brown, T. A. (2011). The structure of personality disorders in individuals with posttraumatic stress disorder. *Personality Disorders: Theory, Research, and Treatment*, *2*(4), 261.

Wright, A. G. C. (2017). The current state and future of factor analysis in personality disorder research. *Personality Disorders: Theory, Research, and Treatment*, *8*(1), 14–25. https://doi.org/10.1037/per0000216

Wright, A. G., Hopwood, C. J., Skodol, A. E., & Morey, L. C. (2016). Longitudinal validation of general and specific structural features of personality pathology. *Journal of Abnormal Psychology*, *125*(8), 1120.

Part 3: Critical Appraisal

Introduction

The research presented here pertains to the question of how psychopathology ought to be classified, which is of great conceptual and practical importance. Proponents of the empirical nosological paradigm explored here suggest that empirical investigations of patterns of comorbidity provide a better characterisation of psychopathology than the diagnostic system, which has historically been informed by clinical opinion. This paradigm is influential and may have a significant impact on clinical care; for example, it has been the driver of the new Diagnostic and Statistical Manual of Mental Disorders (DSM-5) model of personality disorders (Krueger & Markon, 2014) and a recently-formed consortium of researchers in this field have ambitious aims to reconceptualise psychopathology (Kotov et al., 2017). However, this research is pure rather than applied, it requires access to large amounts of data and the factor analytic modelling methods used are complex. This paradigm therefore presents an interesting picture, whereby research which is highly clinically relevant diverges from the practical means by which most working clinicians might undertake it.

Embarking on these research projects alongside clinical training therefore raised some thought-provoking questions, including about the ways in which clinical experience and research skills might be complementary, how research and clinical work influence one another and about the conditions required for individual clinical psychologists to be both 'producers' and 'consumers' of research. This critical appraisal will describe the process of conducting the systematic review and empirical project, reflecting on some of the themes raised and these questions of professional identity.

The systematic review

The systematic review presented here was motivated by two factors. Firstly, I was keen to gain a better understand the *p*-factor, a recent finding which appears to present a significant challenge to the diagnostic system. My attitude on starting the review was not one of scepticism about the *p*-factor, rather I was intrigued by several questions which it raised. These included Caspi and colleagues' (2014) structural hypothesis of psychopathology,

which coheres with other evidence (Meyer et al., 2005) but offers an alternative way of thinking about how internalising, externalising and thought disorder presentations might be related. A second motivation was an interest in the theoretical foundations of research. Systematic reviewing techniques are becoming increasingly sophisticated, with technology now available which can synthesise quantitative data computationally (Thomas, Brunton, & Graziosi, 2010); however, the interpretive components of quantitative work, which are central to the scientific process (Ladyman, 2001), are less often a focus of secondary research. It seemed that this was a particularly important issue in relation to an area with potentially radical implications.

The impression I gained through scoping the review project was that, whereas theoretical accounts in mental health research might draw on empirical work, systematic and interpretive reviews of quantitative research are rare. Therefore, although interpretive research necessarily reflects the view of the researcher, I was keen to use a defined analytical framework and systematic methods wherever appropriate. Adapting a research method designed for qualitative primary research (Braun & Clarke, 2006) so that it could be used to systematically evaluate statistical studies was, to some extent, an iterative process. In practice, the ambitious nature of this project meant that the scope of both the integrative and interpretive syntheses were restricted, and a challenging part of the work was balancing the presentation of enough integrative information to make the critical synthesis meaningful. This necessitated some compromises; for example, although I explored symptom factor loadings and the external validity of the *p*-factor in my empirical study, it would not have been feasible to review these differences between the studies within the bounds of the integrative synthesis. In addition, coding the studies in relation to the pre-determined themes of the interpretive analysis revealed multiple examples of the points made in the review, as well as other potentially relevant issues for discussion, and keeping the thematic analysis discussion succinct and focussed on the points which seemed to be most relevant was challenging.

The project constraints and the chosen review method made reviewing a large number of papers impractical, so the inclusion criteria were designed to be restrictive. At the start of the reviewing process I had not fully appreciated that factor analytic research into Axis I and personality disorders rarely overlapped. However, after my included studies had been identified I realised that the decision to restrict inclusion to papers reporting confirmatory factor analysis (CFA) methods and citing the target paper (Caspi et al., 2014) effectively excluded most studies modelling personality disorder symptomatology, except one study which did not include borderline symptoms (Hoertel et al., 2015). This meant that that my aim of evaluating the theoretical basis of a particular body of research had come at the expense of my aim to better understand the *p*-factor. That the *p*-factor research I was reviewing was missing important psychopathological presentations also influenced the distinction I made in the review between 'the *p*-factor' (which I defined as a general tendency towards comorbidity, identified across different studies) and individual '*p*-statistics' (reflecting study-specific factors, including symptoms measured, population and modelling method).

During the thematic analysis process I became sceptical about some of the assumptions made by several of the study authors. This was partly due to conceptual concerns that the criticisms directed at the diagnostic system which had motivated the empirical nosological approach had not been adequately resolved. For example, the polythetic criteria by which diagnoses are defined seemed to also be implied by any attempts to operationalise latent factors at the level of the individual, perhaps due to a tension between the object of interest in psychology (in this case, 'psychopathology') and observable signs (in this case, 'symptoms') (Essex & Smythe, 1999). However, several of my concerns were specifically informed by consideration of conceptual issues in light of my clinical experience.

Whether grouping people according to symptoms provides the best route to knowledge of psychopathological mechanisms is an open empirical question (Stephan et al., 2016). However, irrespective of this question, my clinical experience is that cross-sectionally

measured symptoms only reflect part of the reality of mental ill-health. This left me sceptical of moves by some of the study authors to talk of redefining qualitative notions such as ‘severity’ or ‘degree of pathology’ in terms of statistical results alone. Relatedly, I was unsure whether the conclusion that symptom-focussed transdiagnostic treatments were indicated by the *p*-factor reflected my own clinical experience. On the one hand, as some of the researchers in this area acknowledge (Noordhof et al., 2015; Wright, 2017), general etiological factors do not necessarily imply that interventions should not focus on particularities of different presentations. Equally, my experience was that other dimensions of psychopathology, such as relational difficulties, could be particularly relevant in the treatment of people with more complex presentations. Just as I was unconvinced by the formulation of ‘severity’ simply in terms of comorbidity of symptoms, I was unsure whether conceptualising ‘transdiagnostic treatment’ in terms of a flexible approach to symptom-relief alone was clinically indicated.

These concerns seemed pertinent in relation to models of balancing clinical work and research within clinical psychology, specifically reflecting the risk that the ‘clinical scientist’ could become removed from the realities of clinical work (Barker, Pistrang, & Elliott, 2002). This type of research, by virtue of its abstract nature, might be particularly likely to attract those interested in science over practice, and the practicalities of developing the necessary research skills and gaining access to data could create further barriers to these questions being investigated by people with mixed clinical and research careers. However, given the above observations, I wondered whether a clinically-informed view might be particularly important in the interpretation of ‘pure’ research, as its abstract quality might encourage an intellectualised and technical approach.

The empirical study

Data collection and designing the questions

The empirical study presented here used data collected from a larger project, the Probing Social Exchanges Study. The broader aims of this study are to understand processes relevant to social cognition in personality disorder and the study is currently being extended to depression. The research questions are around mentalising and other relational processes, with mechanistic research questions being answered within a computational psychiatry paradigm (for example, King-Casas et al., 2008). I was keen to join this study because it offered the opportunity to gain skills and experience in a range of assessment techniques, beyond what would be typical in a clinical context, including behavioural games, diagnostic interviewing and the Adult Attachment Interview. The training and process of data collection reinforced my thoughts that cross-sectionally measured symptoms only constitute a partial description of people's functioning and experience, and that it was important not to reify such a narrow part of the clinical picture.

However, notwithstanding my reservations about the limitations of the empirical nosology, joining the Probing Social Exchanges Study offered a rare possibility of modelling a *p*-statistic reflecting a wide range of symptoms in a sample largely comprised of people with personality disorders. The systematic review had a significant influence on my choice of hypotheses, both with regard to the model structures tested and the relationship between the *p*-factor and child maltreatment (Brodbeck et al., 2014; Caspi et al., 2014; Kim & Eaton, 2015; Laceulle et al., 2015). Several papers offering theoretical accounts of the *p*-factor were identified during the systematic search, although they did not meet the review inclusion criteria. These included the hypothesis that the *p*-factor represented a lack of openness to social information in the context of impairments in mentalising (Fonagy & Campbell, 2015), which formed the basis of an additional hypothesis.

Clinically-informed interpretation

As discussed, in the course of completing the systematic review I had become aware of the importance of keeping clinical and theoretical issues in mind when interpreting empirical findings, and I aimed to do this in my own study, albeit with an ever-increasing awareness that it is easier to criticise than to do things differently. I was particularly determined to discuss the empirical findings in terms of observed patterns of comorbidity, without making a firm commitment to 'real' latent entities or the 'natural structure' of psychopathology. In addition, I aimed to avoid reification through exploring particular statistical relationships, both across the models tested and with reference to other studies where possible, whilst emphasising the contingencies these statistical relationships might reflect.

Interpreting the results in a clinically-informed way was important in relation to the trend observed in the fit of the models tested, whereby the modified bifactor model with no internalising group factor fitted better than the models without thought disorder and borderline group factors. In addition, on statistical grounds the model without an internalising factor could be considered preferable to the full bifactor model, due to the insignificant factor loadings in the full bifactor. I discussing this with my supervisors, who emphasised the importance of not denying a clinically-meaningful aspect of reality on the basis of statistics, prompting me to think more about what might have influenced these results. This led me to think more about what CFA models do describe (that is, comorbidity) and what they do not reflect; factors including prevalence, general impairment and (in some cases) duration of presentation. On this basis I re-examined possible reasons why two reviewed studies found that thought disorder symptoms did not form a group factor (Caspi et al., 2014; Laceulle et al., 2015), and wondered whether a possible reason for this was that these studies were the only ones to model longitudinal data. For the reasons outlined in the empirical paper and briefly summarised here, there are a number of potential reasons for the results obtained in this and other studies, some of which may be spurious, and this interpretation may not be

correct. However, this was an instance where I realised that clinical judgement might not only temper the particular conclusions of a study but – importantly – that it might also be able to offer valuable feedback on the theoretical implications of quantitative methods.

Clinically-generated ideas and the research-practice cycle

Proponents of the empirical nosology have suggested that psychopathology research might be approached from both the perspectives of nosology and the psychological processes underpinning symptoms; for example, the recently-formed empirical nosology consortium suggested that their research might 'meet' the Research Domain Criterion (RDoC) programme in this way (Kotov et al., 2017). RDoC is narrow insofar as its stated aim is to investigate only psychological processes which are underpinned by identifiable neural substrates (Insel et al., 2010). However, this notion of approaching psychopathology from 'two sides' fitted with the aims of my empirical project, which looked at the relationship between symptomatology, childhood maltreatment and reflective function. However, the theories of latent vulnerability and mentalising have implications beyond just symptoms and comorbidity, including for developmental processes and additional qualitative dimensions of psychopathology. For example, functional neuroimaging evidence shows similar patterns are observed in victims of maltreatment, whether or not they are currently symptomatic (McCrory et al., 2017), and impairments in reflective function are associated with psychosocial impairment, quality of life and well-being (Fonagy et al., 2016). Therefore an important question would seem to be whether a narrow nosology based on cross-sectional symptoms might run the risk of obscuring as well as elucidating.

Reflecting on the synergies between research and clinical practice led to me think about the origins of theory, particularly in relation to the concept of mentalising, which was formulated in response to clinical observations within a therapeutic context (Fonagy, 1991). The foundations of mentalising theory have since been built, with methods of measurement developed that have facilitated research (Fonagy et al., 2016). It is a common conception that clinicians are consumers of research, but they may in fact be more likely to generate

innovative research ideas (Stiles, 1992). In the case of mentalising, whilst research has been important in establishing its relevance, it seems inconceivable that the theory could have been developed except through clinical insight. My own experience is that, whilst it may be reasonably intuitive that most people adopt an 'intentional stance' towards themselves and others, clinical practice provides a much fuller understanding of the complex concept of mentalising, and I often find myself surprised by the idiosyncratic ways in which difficulties with mentalising can be associated with particular sources of distress and symptomatic presentations.

In this critical appraisal I have focussed on some areas of scepticism in relation to the empirical nosology. However, despite its potential limitations as the sole means of classifying psychopathology and as a generator of theories, empirically-grounded comorbidity models also have strengths. The *p*-factor research in particular, through emphasising shared variance, offers a way of analysing general effects of risk factors and other psychological processes on psychopathology, thus supporting new insights through deprioritising particular symptoms. To again use the example of mentalising; mentalising theory, which originated from clinical observation and which has been observed within clinical contexts to be relevant to various presentations, has been developed to the point at which the capacity to mentalise can be measured within large samples; and in turn the investigation of mentalising in relation to the empirical nosology, presented here, provides empirical support for its broad relevance. This illustrates the 'cycle' of research (Barker et al., 2002), where ideas which may be inspired by clinical insight are exposed to "scientific quality control" (Stiles, 1992; p. 306).

Balancing clinical and research work in practice

In this critical appraisal I have discussed aspects of my own experience of the research process which highlight some drawbacks of a model where 'clinical scientists' generate knowledge which clinicians consume. However, although the 'scientist-practitioner' model of the clinical psychologist as someone skilled in both research and clinical work may be an ideal, in practice such a balance can be difficult to attain (Barker et al., 2002); although

the clinical psychology doctorate allows protected research time, split roles for qualified psychologists are rare. The type of research I chose to embark on may not be the choice of all of my colleagues and its clinical relevance is unlikely to be direct. Therefore, as I near the end of training, I wondered whether there were further questions generated by my study and research into the *p*-factor generally, which might be more feasible to conduct in an applied framework or alongside clinical practice. Here I have drawn on the models of the 'applied scientist' and 'local scientist', which prioritise practice-based evidence and qualitative and small *N* designs, and the 'evidence-based practitioner' model, which prioritises basing clinical decisions on evidence (Barker et al., 2002).

The *p*-factor might be a helpful way to think about complex patients, who may be more likely to move between services or fall between the gaps of current service provision. Important questions which could be addressed within a practice-based evidence framework include how psychological processes such as mentalising or those affected by childhood maltreatment might influence outcomes of evidence-based treatments. In addition, latent factor research has identified distinct patient groups using data collected routinely within Improving Access to Psychological Therapies services, and found that certain groups have significantly poorer treatment outcomes in that service context (Saunders, Cape, Fearon, & Pilling, 2016). Although comorbidity was not measured as part of this research, the groups with worse outcomes tended to have more severe symptoms, be older (perhaps indicating longer-term presentations) and to be on benefits (which is associated with the *p*-factor; Caspi et al., 2014), indicating they might be high-*p* individuals. Research investigating links between the empirical nosology and routinely-collected data which could serve as 'indicators' of high-*p* status could be a helpful step in understanding more about the clinical implications of *p*.

Researchers working in the area of empirical nosology acknowledge that general functioning is not reflected in their taxonomy, and tentatively suggest 'diagnosis' might move towards a model similar to that used to diagnose learning disabilities, in which the adaptive

functioning is taken into account as well as intelligence (Kotov et al., 2017). Nosological models with multiple axes include psychodynamic diagnoses, which take into account different dimensions of functioning (Gordon, 2010), and the model of personality disorders for further investigation in Section III of the DSM-5 (American Psychiatric Association, 2013), for which the criteria include general impairments in self or interpersonal functioning as well as specific traits. Clinicians would be well-placed to identify or further describe other clinically-relevant dimensions of functioning which might interact with an empirical nosology. For example, although not examined here, it has been suggested that epistemic (mis)trust in the context of impairments in mentalising might underpin psychopathology and influence outcomes of psychological treatment (Fonagy & Allison, 2014). This is a broad question which could lend itself to various methods, including research focussed on qualitative aspects of the therapeutic process.

Conclusion

In conducting this research I had a rare opportunity to join a large, established study, and thereby to answer research questions I would not ordinarily have been able to ask in the course of a D.Clin.Psy project. I became familiar with a particular paradigm from both a theoretical perspective, through conducting an interpretive review, and from the perspective of an empirical researcher, as I also needed to interpret my findings and make methodological compromises. I believe that the necessity of striking a balance between engaging with the empirical nosology and retaining a critical distance improved both research projects, as I was mindful of how things seemed from the 'other side'. Finally, I was fortunate to have supervision which helped me think about the research from both an academic and clinical perspective, again, emphasising the importance of reflection in any research relating to mental health, including – and perhaps especially – in relation to pure research.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.

Barker, C., Pistrang, N., & Elliott, R. (2002). *Research Methods in Clinical Psychology: An Introduction for Students and Practitioners* (2nd ed.). Chichester: John Wiley & Sons. Retrieved from http://www.amazon.co.uk/Research-Methods-Clinical-Psychology-Practitioners/dp/047149089X

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brodbeck, J., Stulz, N., Itten, S., Regli, D., Znoj, H., & Caspar, F. (2014). The structure of psychopathological symptoms and the associations with DSM-diagnoses in treatment seeking individuals. *Comprehensive Psychiatry*, *55*(3), 714–726.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., … Poulton, R. (2014). The p factor one general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, *2*(2), 119–137.

Essex, C., & Smythe, W. E. (1999). Between Numbers and Notions A Critique of Psychological Measurement. *Theory & Psychology*, *9*(6), 739–767.

Fonagy, P. (1991). Thinking about Thinking: Some Clinical and Theoretical Considerations in the Treatment of a Borderline Patient. *Int. J. Psycho-Anal.*, *72*, 639–656.

Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy*, *51*(3), 372.

Fonagy, P., & Campbell, C. (2015). Bad Blood Revisited: Attachment and Psychoanalysis, 2015. *British Journal of Psychotherapy*, *31*(2), 229–250. https://doi.org/10.1111/bjp.12150

Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y.-W., Warren, F., Howard, S., … Lowyck, B. (2016). Development and Validation of a Self-Report Measure of Mentalizing: The

Reflective Functioning Questionnaire. *PLOS ONE*, *11*(7), e0158678. https://doi.org/10.1371/journal.pone.0158678

Gordon, R. M. (2010). Psychodynamic Diagnostic Manual. *Corsini Encyclopedia of Psychology*.

Hoertel, N., Franco, S., Wall, M. M., Oquendo, M. A., Kerridge, B. T., Limosin, F., & Blanco, C. (2015). Mental disorders and risk of suicide attempt: a national prospective study. *Molecular Psychiatry*, *20*(6), 718–726.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Kozak, M., Pine, D. S., … Wang, P. (2010). Research Domain Criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiat*, *167*, 748–751.

Kim, H., & Eaton, N. R. (2015). The hierarchical structure of common mental disorders: Connecting multiple levels of comorbidity, bifactor models, and predictive validity. *Journal of Abnormal Psychology*, *124*(4), 1064.

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science*, *321*(5890), 806–810.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., … Clark, L. A. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454.

Krueger, R. F., & Markon, K. E. (2014). The role of the DSM-5 personality trait model in moving toward a quantitative and empirically based approach to classifying personality and psychopathology. *Annual Review of Clinical Psychology*, *10*, 477–501. https://doi.org/10.1146/annurev-clinpsy-032813-153732

Laceulle, O. M., Vollebergh, W. A., & Ormel, J. (2015). The Structure of Psychopathology in Adolescence Replication of a General Psychopathology Factor in the TRAILS Study. *Clinical Psychological Science*, 2167702614560750.

Ladyman, J. (2001). *Understanding Philosophy of Science* (1st Edition edition). London ; New York: Routledge.

McCrory, E. J., Gerin, M. I., & Viding, E. (2017). Annual Research Review: Childhood maltreatment, latent vulnerability and the shift to preventative psychiatry–the contribution of functional brain imaging. *Journal of Child Psychology and Psychiatry*, *58*(4), 338–357.

Meyer, S. E., Bearden, C. E., Lux, S. R., Gordon, J. L., Johnson, J. K., O'Brien, M. P., … Cannon, T. D. (2005). The psychosis prodrome in adolescent patients viewed through the lens of DSM-IV. *Journal of Child & Adolescent Psychopharmacology*, *15*(3), 434–451.

Noordhof, A., Krueger, R. F., Ormel, J., Oldehinkel, A. J., & Hartman, C. A. (2015). Integrating autism-related symptoms into the dimensional internalizing and externalizing model of psychopathology. The TRAILS Study. *Journal of Abnormal Child Psychology*, *43*(3), 577–587.

Saunders, R., Cape, J., Fearon, P., & Pilling, S. (2016). Predicting treatment outcome in psychological treatment services by identifying latent profiles of patients. *Journal of Affective Disorders*, *197*, 107–115. https://doi.org/10.1016/j.jad.2016.03.011

Stephan, K. E., Bach, D. R., Fletcher, P. C., Flint, J., Frank, M. J., Friston, K. J., … Breakspear, M. (2016). Charting the landscape of priority problems in psychiatry, part 1: classification and diagnosis. *The Lancet Psychiatry*, *3*(1), 77–83. https://doi.org/10.1016/S2215-0366(15)00361-2

Stiles, W. B. (1992). Producers and consumers of psychotherapy research ideas. *The Journal of Psychotherapy Practice and Research*, *1*(4), 305.

Thomas, J., Brunton, J., & Graziosi, S. (2010). *EPPI-Reviewer 4: software for research synthesis. EPPI-Centre Software*. London: Social Science Research Unit, UCL Institute of Education.

Wright, A. G. C. (2017). The current state and future of factor analysis in personality disorder research. *Personality Disorders: Theory, Research, and Treatment*, *8*(1), 14–25. https://doi.org/10.1037/per0000216

Appendices

Appendix A1: List of acronyms used in the systematic review

Statistical terms

General

CI: confidence interval

sd: standard deviation

Statistical modelling terms

CFA: confirmatory factor analysis

EFA: exploratory factor analysis

SEM: structural equation model/modelling

Fit indices

CFI: comparative fit index

RMSEA: root mean square error of approximation

SRMR: standardised root mean square residual

TLI: Tucker-Lewis index

Studies

ALSPAC: Avon Longitudinal Study of Parents and Children, a prospective cohort study of children born 1991-1992; Avon, UK

CAP: Climate Schools and Preventure, an RCT testing an intervention for substance misuse in adolescence; Sydney and Melbourne, Australia

IMAGEN: a cohort study investigating neurocognitive factors associated with psychopathology in adolescents; Ireland, Germany, France, UK

NESARC: National Epidemiologic Survey on Alcohol and Related Conditions, a cross-sectional population study of alcohol misuse and other psychopathology; US

ROOTS: prospective cohort study investigating factors associated the development of psychopathology in adolescents; UK

TRAILS: TRacking Adolescents' Individual Lives Survey, a prospective population and clinical cohort study of psychological and physical health in adolescence; Groningen, The Netherlands

Models

Latent factors

AGG: aggression

ANX: anxiety

ATT-OR: attention-orientation problems

ASD: autism spectrum

DEP: depression

DIST: distress

FEAR: fear

INFO: information processing

IS: interpersonal sensitivity

INT: internalising

EXT: externalising

NERV: nervous tension

PHOB: phobic fear

SOM: somatic problems

SUI: suicidal ideation

Diagnoses (designated by upper case lettering)

ADHD: attention deficit hyperactivity disorder

ALC: alcohol dependence

ANX: anxiety

ASPD: Antisocial Personality Disorder

APD: Avoidant Personality Disorder

BD: bipolar disorder or mania

CAN: cannabis addiction

CD: conduct disorder

DEP: depression

DPD: Dependent Personality Disorder

DRUG: drug addiction (hard drugs)

DYST: dysthymia

ED: eating disorder

FEAR: fear

GAD: generalized anxiety disorder

GAMB: gambling addiction

HPD: Histrionic Personality Disorder

MDD: major depression

OCD: obsessive compulsive disorder

OCPD: obsessive compulsive personality disorder

PAN: panic disorder

P/PAN: phobia/panic disorder

PHOB: specific phobia

PPD: Paranoid Personality Disorder

SAD: social anxiety disorder

SEP: separation anxiety

SCHIZ: schizophrenia

TOB: tobacco addiction

Symptoms (designated by lower case lettering)

a/dep: anxious depression

agor: agoraphobia

as-b: behaviour and emotions not tuned to social situation (autistic spectrum)

as-st: stereotyped behaviour (autistic spectrum)

as-o: orientation-problems in time, place, or activity (autistic spectrum)

as-s: reduced contact and social interests (autistic spectrum)

as-r: resistance to change (autistic spectrum)

as-u: difficulties in understanding social information (autistic spectrum)

asoc: antisocial behaviour/ delinquency

att: attentional difficulties

cd: conduct disorder

del: delusions

fear: combined symptoms of phobia, SAD and agoraphobia

hall: hallucinations

imp: hyperactivity/impulsivity

is: interpersonal sensitivity

opp: oppositional defiant

pan/som: panic and somatic symptoms

psy: psychotic experiences

rule: rule-breaking

som: somatic symptoms

sad: social anxiety

sch: school phobia

sep: separation anxiety

td: thought disorder

w/dep: withdrawn depression

Appendix A2: Fit indices for best-fitting models

Parsimony statistics are not quoted as these are only meaningful as a means of comparing models; STOCHL2015 is not included in the table as models are only assessed using parsimony statistics in this study.

Study	Model	Parameters	Chi-sq	df	CFI	TLI	RMSEA [90% CI]	SRMR
Child/ adolescent population								
CARRAGHER2016	Bifactor	135	1245.193	855	.98	.97	.014 [.013 – .016]	
CASTELLANOS2016	Bifactor		175.98	42	.94		.038	.022
LACEULLE2015	Bifactor (no TD)		4665.65	716	.90	.89	.050 [.048 – .051]	
LAHEY2015	Bifactor		192.11	9	.97		.061 [.054 – .068]	
MARTEL2016	Bifactor*	84	174.729		.98	.97		
NOORDHOF2015	Bifactor (T2)				.98		.05	
PATALAY2015	Bifactor				.95	.94	.05	
WALDMAN2016	Bifactor**		424	34		.96	.060	.026
Adult population								
BRODBECK 2014	Group factor	264	3432	1276	.95	.93	.041 [.039 – .043]	
	Bifactor	267	4307	1270	.92	.92	.049 [.047 – .050]	
CASPI2014	Group factor		1,737.159	1018	.96	.96	.027 [.024 – .029]	
	Bifactor (no TD)		1,652.586	1012	.97	.96	.025 [.023 – .027]	
HOERTEL2015	Bifactor+				.98	.98	.011	
KIM2015	Bifactor				.99	.99	.012	
SUBICA2015	Bifactor		772.005	88	.98	.974	.090 [.084 – .096]	

Key: * fear and distress as orthogonal factors; ** MDD and GAD only load on *p* and not INT; + general population model

Appendix B1: List of acronyms used in the empirical paper

Measures

APSD: Antisocial process screening device

BSI: Brief symptom inventory

CTQ: Child Trauma Questionnaire

DERS: Difficulties in Emotion Regulation Strategies Scale

RFQ: Reflective Function Questionnaire

RFQ_C: RFQ certainty about mental states subscale

RFQ_U: RFQ uncertainty about mental states subscale

GPTS: Green's paranoid thought scale

LHA: Life History of Aggression scale

PAI-BOR: Personality Assessment Inventory-Borderline Features

PCL-S: PTSD checklist - specific

SPQ: Schizotypal Personality Questionnaire

Statistical terms

General

CI: confidence interval

Modelling

CFA: confirmatory factor analysis

EFA: exploratory factor analysis

FIML: full information maximum likelihood

MLR: robust maximum likelihood

Fit indices and parsimony statistics

AIC: Akaike’s Information Criterion

ABIC: Adjusted Bayesian Information Criterion (adjusted for sample size)

CFI: comparative fit index

RMSEA: root mean square error of approximation

SRMR: standardised root mean square residual

TLI: Tucker–Lewis index

Models: latent factors

ASOC: antisocial

INT: internalising

TD: thought disorder

BOR: borderline

Appendix B2: Research ethics committee approval

Research Ethics Committee (REC) for Wales
Sixth Floor, Churchill House
17 Churchill Way
Cardiff CF10 2TW
Telephone : 029 2037 6829
Fax : 029 2037 6824

E-mail : corinne.scott@wales.nhs.uk

Website : www.nres.nhs.uk

09 October 2012

Professor Peter Fonagy
HoD, Department of Clinical, Educational and Health Psychology, UCL
UCL
Gower Street
London WC1N 3BG

Dear Professor Fonagy

Study title: **Probing Social Exchanges – A Computational Neuroscience Approach to the Understanding of Borderline and Anti-Social Personality Disorder**
REC reference: **12/WA/0283**

Thank you for your letter of 25 September 2012, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information was considered by a sub-committee of the REC at a meeting held on 05 October 2012. A list of the sub-committee members is attached.

Confirmation of ethical opinion
On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Ethical review of research sites
NHS sites
The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Conditions of the favourable opinion
The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission ("R&D approval") should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements.

Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the

R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of approvals from host organisations

- *The Clinical / Probation Service information sheet, page two paragraph one, has the phrase "which is a psychiatric interview" twice; one of these instances should be removed;*
- *The word "However" should be removed from the start of the first paragraph of page three under "What are the possible disadvantages and risks of taking part?";*
- *The second paragraph of the same section is the same sentence repeated twice, and one of these instances should be removed;*
- *The Healthy volunteers information page three, the word "However" should be removed from the start of the first paragraph of page three under "What are the possible disadvantages and risks of taking part?"*

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

You should notify the REC in writing once all conditions have been met (except for site approvals from host organisations) and provide copies of any revised documentation with updated version numbers. Confirmation should also be provided to host organisations together with relevant documentation.

Approved documents
The final list of documents reviewed and approved by the Committee is as follows:

Document	Version	Date
Advertisement	Letter of invitation = advertisement material as well; version 1.1	22 August 2012
Covering Letter	signed Tobias Nolte, Anna Freud Centre	22 August 2012
Evidence of insurance or indemnity	Arthur J Gallagher International certificate of insurance - University College London - expires 01 August 2013	30 July 2012
GP/Consultant Information Sheets	1	22 August 2012
Investigator CV	Professor Fonagy; version 1.1	22 August 2012
Investigator CV	Dr Feigenbaum; version 1.1	22 August 2012
Investigator CV	Tobias Nolte; version 1.1	22 August 2012
Investigator CV	P Read Montague; no version or date	
Letter from Sponsor	signed David Wilson, University College London	21 August 2012
Letter of invitation to participant		22 August 2012
Other: Risk and Safety Protocol	1.1	22 August 2012
Other: Data Protection Form	no version or date	
Other: Additional details regarding MRI data	1.1	22 August 2012
Other: Consent to contact form	1.1	22 August 2012
Participant Consent Form: Healthy volunteers	1.2	
Participant Consent Form: Clinical / Probation service	1.2	
Participant Information Sheet: Genetics	1.1	22 August 2012
Participant Information Sheet: Healthy volunteers	1.2	

Appendix B3: Participant information sheet

Understanding the Social Brain in Healthy Volunteers and People with Psychological Difficulties.

This study has been approved by the Research Ethics Committee for Wales (Project ID Number): 12/WA/0283.

We would like to invite you to participate in this research project.

You are being invited to take part in a research study. You should only participate if you want to. Before you decide whether to take part, this sheet will give you some more information about why the study is being carried out, what you would be asked to do if you decide to take part, and how the study will be conducted. Please take some time to read this sheet, and to discuss it with other people if you wish. You are also very welcome to ask any further questions about the study, or if you find anything on this sheet unclear.

Why is this study being done?

With the proposed project we plan to investigate the brain activation patterns of people suffering from personality disorders (both in adults and adolescents) or similar traits and compare them with healthy control participants. Only little is known about the neurobiology of Borderline and Antisocial Personality Disorders. Our study design will address some of these. This will hopefully allow us to gain a better understanding of the disorders and to develop more informed and effective treatments from which clients will benefit.

Why have you been invited to take part?

You have been invited to take part in the study because you have recently been assessed by a clinician at one of the clinical or probation services currently collaboration with the research team.

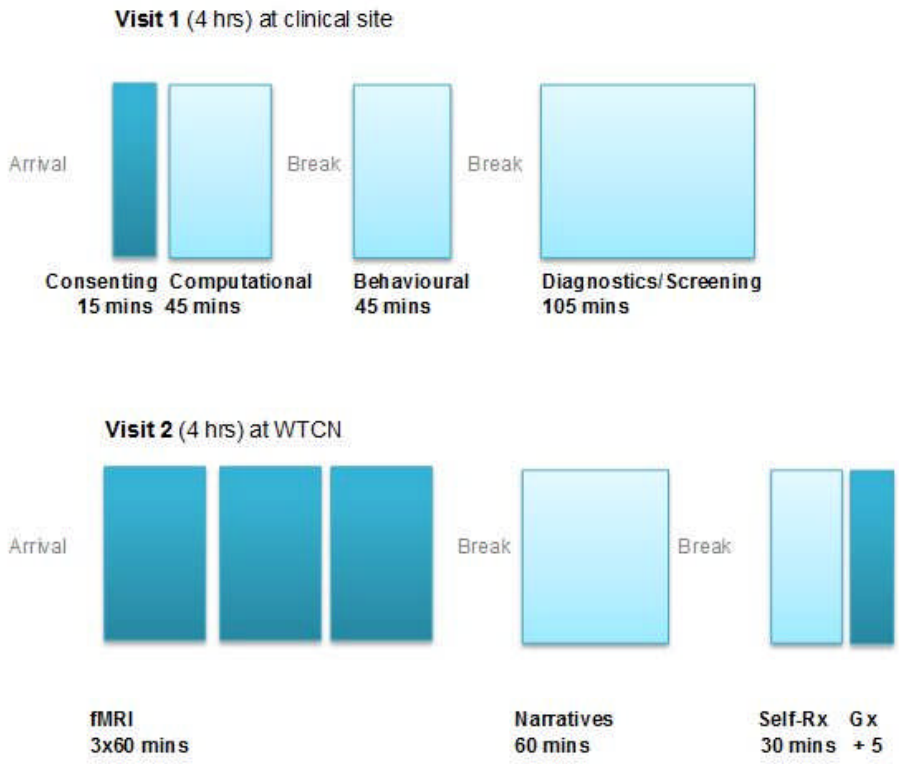
Do I have to take part?

No. Taking part in the study is entirely voluntary. It is your choice whether or not you would like to participate. Deciding not to take part in the study will not affect the care you receive from services either now or in the future. If you do decide to participate, you will be given this information sheet to keep, and you will later be asked to sign a consent form stating that you wish to take part. If you do give consent to take part in the study, you are still free to leave the study at any point, without giving a reason. This will not affect the care you are currently receiving, or will receive in the future. If you leave, any information that we have already collected from you will be destroyed.

What will happen if I decide to take part?

If you wish to take part in the study, then you can get in touch with the research team or provide your contact details so that we can arrange a time to discuss the study in more detail and to book in the assessments if consent is obtained. We can then contact you to arrange a convenient time to meet. At this meeting you will meet a member of the research team and you can ask any other questions you may have. You will then be asked to sign a consent form to say that you wish to take part in the study. You will also be asked about your eligibility for brains scans as not every person can undergo these.

Study overview:



There will be two or three assessments with approximately 8 hours in total duration. In the first assessment, which will be held at the clinical site or the probation service, you will be asked to fill in questionnaires on personality functioning, developmental history, symptomatology etc. You will then perform several computer-based cognitive tasks and have a SCID I and II (relevant sections only) which is a psychiatric interview that takes approximately 30 to 60 minutes to complete. Any of these measures that have already been routinely obtained at your service will not be repeated if you are happy for your service to share the data with us (your consent provided).

If you agree to participate in this study you will be asked to come to the Wellcome Trust Centre for Neuroimaging on one occasion. The experiment will consist of 5 computerised tasks (which you will do whilst lying in a magnetic resonance imaging (MRI) brain scanner). In the tasks you will have to perform some tasks such as responding to written cues using different buttons to estimate or compare different events or conditions (similar to simple computer games) In some of them you will play another person who is being scanned at a different laboratory at the Principal Investigator's second laboratory at Virginia Tech University.. This phase will last roughly 3 hours but it is broken down into 3 sections of 60 minutes maximum with lots of breaks. After each hour you will have a longer break and leave the scanner. Most people find the tests quite straightforward and

interesting to do. After the scanning, we will ask you to answer some further questions regarding the same or similar events or conditions, fill out several questionnaires and you will be administered an interview regarding experiences in your childhood which usually takes another 45 minutes and which will be audio-recorded and transcribed before being coded for attachment by a reliable and experienced member of the research team. Before coding, all identifiable information will be removed from the audio file for anonymity.

If you have a tattoo, we will ask you to participate in a study that investigates any adverse effects which may occur as a result of MRI, such as heating or pulling on the tattoo.

No part of the study is compulsory and there will be separate consent sections for each part of the study.

What is functional magnetic resonance (fMRI) and what are the potential risks?

An MRI scanner takes pictures of your brain and measures the activity of different parts of it. The MRI scan procedure is painless and safe – these procedures are done hundreds of times a day all over the world. However, the MRI scanner makes loud noises while it is operating; we will provide you with headphones or earplugs to reduce the noise to safe levels. Some people find being in an MRI scanner makes them feel anxious and/or claustrophobic, even if they have not experienced claustrophobia before. A member of staff will be in constant contact with you via the intercom, and if you feel uncomfortable in any way the scanning can be stopped. Before you get into the MRI scanner the person who operates the scanner will explain the procedure to you and answer your questions. There is no radiation involved. MRI scans work using very strong magnetic fields. Therefore it would be dangerous for anyone with any magnetic metal in their body to go near the scanner, since that metal might move towards the magnet. You will not be able to participate in the MRI scan if you do have such metal in your body. Examples include: pace-makers; piercings; certain tattoos (which are sometime made with metallic inks) and screws from surgery. Fillings are not magnetic and are therefore not a problem. **If you are not sure whether you are able to participate in the MRI scan due to the presence metal in your body, please ask a researcher.**

What are the possible disadvantages and risks of taking part?

We will support you if you become upset. A specific Risk and Safety protocol for this study has been developed. You will be given time at the end of the study to be fully debriefed with a member of the research team and provided with a handout on emotional regulation skills, and crisis phone numbers and details of clinical services to contact. Your personal therapist or probation officer will also be aware of your participation in the study and able to support you should you find discussing your experiences difficult. Should you feel overwhelmed or acutely distressed during or at the end of the assessments, we you will be appropriately looked after by an experienced clinician.

Some people find the experience of being in the brain scanner uncomfortable or distressing as it is very noisy in you will have to lie still for a long time in a narrow tube.

Should any abnormalities be found during the scan a qualified Neurologist will be asked to review the image and if necessary contact your GP regarding any concerns.

What are the possible benefits of taking part?

You may find it interesting to complete these tasks and the information gathered during this study will also help to inform our understanding of treatment for Personality Disorders, which will hopefully be a step towards helping improve interventions in the future.

Will I be paid for taking part in the study?

As an acknowledgement of your time, we will be offering you a flat rate of £10 per hour for your participation with additional compensation depending on your performance on some of the tasks. If you agree to give a saliva and blood sample, we will be offering you an additional £30.

Who will know you are taking part in the study?

We will inform your personal therapist or probation officer if you have been recruited via these services. We will inform your GP of your participation in this study, but information collected during all stages of the study will be kept strictly confidential. All information will only be viewed by members of the research teams at University College London and Virginia Tech University in the US.. However, if through the course of the study it was found that you are at immediate risk of harm to yourself or others, this information will be shared with your therapist or GP and, if necessary, emergency services.

Your consent form will be kept in a separate location from all your other data, ensuring that this remains anonymous. All data will be stored in secure locations whereby a participant ID will be assigned to your data, not identifiable personal information and the results of your tasks will be recorded on computers or flash drives which are password protected. Any published data will also be entirely anonymous meaning individuals cannot be identified.

Some of the MRI data will be transferred for analysis to the Principal Investigator's second laboratory at Virginia Tech University in the US. Those data will be anonymised and no identifiable personal information will be shared or transferred.

The data from this study will be stored in accordance with the UCL and NHS Data Protection and Records Management policies.

All data will be collected and stored in accordance with the Data Protection Act 1998.

What will happen to the results of the research study?

The results will be written up in the form of reports to be submitted to scientific journals or presented at conferences. As mentioned, you will not be identifiable from these results. On completion and if you request it you will be sent a report of the study.

What if there is a problem?

Every care will be taken in the course of this study. However, in the unlikely event that you are injured by taking part, compensation may be available.

If you suspect that the injury is the result of the Sponsor's (University College London) negligence then you may be able to claim compensation. After discussing with your research doctor, please make the claim in writing to Dr. Janet Feigenbaum or Dr Tobias Nolte on behalf of the Chief Investigators (Profs Read Montague and Peter Fonagy) who are based at University College London. The Chief Investigator will then pass the claim to the Sponsor's Insurers, via the Sponsor's office. You may have to bear the costs of the legal action initially, and you should consult a lawyer about this.

If you wish to complain, or have any concerns about any aspect of the way you have been approached or treated by members of staff you may have experienced due to your participation in the research, National Health Service or UCL complaints mechanisms are available to you. Please ask your research doctor if you would like more information on this. In the unlikely event that you are harmed by taking part in this study, compensation may be available to you. If you suspect that the harm is the result of the Sponsor's (University College London) or the hospital's negligence then you may be able to claim compensation. After discussing with your research doctor, please make the claim in writing to the Prof Fonagy who is the Chief Investigator for the research and is based at UCL, Research Department of Clinical, Educational and Health Psychology, 1-19 Torrington Place, London, WC1E 7HB. The Chief Investigator will then pass the claim to the Sponsor's Insurers, via the Sponsor's office. You may have to bear the costs of the legal action initially, and you should consult a lawyer about this

Who has reviewed this study?

This study has been reviewed by the **REC for Wales 12/WA/0283**

Contact Details

If you wish to contact the research team to discuss any of the information further or any concerns you have about the study, then please do so by getting in touch with the members of the research team listed below:

If you feel that we have not addressed your questions adequately or if you have any concerns about the conduct of the research team, then please contact my supervisor Dr. Janet Feigenbaum (Strategic and Clinical Lead for Personality Disorder Services, North East London NHS Foundation Trust and Senior Lecturer, Research Department of Clinical, Educational and Health Psychology, UCL) on 07957 919 961or by email at janet.feigenbaum@nhs.net.

Janet Feigenbaum, PhD
Research Department of Clinical, Educational and Health Psychology
General Office, Room 436, 4th Floor
1-19 Torrington Place, London, WC1E 7HB

Tobias Nolte MD

Wellcome Trust Centre for Neuroimaging & Research Department of Clinical, Educational and Health Psychology
12 Queen Square
London
WC1N 3BG
Tobias.nolte@annafreud.org

Thank you very much for taking the time to read this information sheet.

Appendix B4: Participant consent form

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

Project Title:

Understanding the Social Brain in Healthy Volunteers and People with Psychological Difficulties.

This study has been approved by the Research Ethics Committee for Wales (Project ID): 12/WA/0283.

Thank you for your interest in taking part in this research. Before you agree to take part, the person organising the research must explain the project to you.

If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you to decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.

Participant's Statement

I

- have read the notes written above and the Information Sheet, and understand what the study involves. I am also aware that I can consent to certain aspects of the study in order to participate in them whereas I can withhold my consent for others parts.
- understand that if I decide at any time that I no longer wish to take part in this project, I can notify the researchers involved and withdraw immediately.
- consent to the processing of my personal information for the purposes of this research study.
- understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.
- understand that some of the MRI data will be transferred for analysis to the Principal Investigator's second laboratory at Virginia Tech University in the USA and will therefore no longer be subject to EEA data protection laws but that this data will be anonymised and no identifiable personal information will be shared or transferred.
- agree that the research project named above has been explained to me to my satisfaction and I agree to take part in this study.
- I agree that my non-personal research data may be used by others for future research. I am assured that the confidentiality of my personal data will be upheld through the removal of identifiers.

- I understand that part of my participation will be audio-recorded (the interviews) and I consent to the anonymous use of this material as part of the project.
- I agree to be contacted in the future by UCL researchers who would like to invite me to participate in follow-up studies.
- I understand that the information I have submitted will be published as a report and that I can request a copy. Confidentiality and anonymity will be maintained and it will not be possible to identify me from any publications.
- I agree that the research team might re-contact me in case that additional data has to be obtained or for follow-up studies.

Please initial the statements below if you agree with them:	**Initial here**
I agree to take part in the general part of the PD-CPA study as outlined in the information Sheet and to all points listed above. (a separate consent for the MRI, tattoo component, and genetics component follows below).	
I agree to the audio recording of interviews and I consent to the anonymous use of this material as part of the project.	
I agree that some of the study data will be shared with the collaborating laboratory at Virginia Tech University in the USA.	
I agree to the audio recording of interviews and I consent to the anonymous use of this material as part of the project.	
I agree that some of the study data will be shared with the collaborating laboratory at Virginia Tech University in the USA.	
I understand that relevant sections of medical and or probation notes and data collected during my clinical assessment and during the study from me, may be looked at by individuals from the PD-CPA research team, my clinician or from the NHS Trust, where it is relevant to our taking part in this research. I give permission for these individuals to have access to my records.	
I agree that the PD-CPA research team can contact me about coming in for up to two follow-up sessions over the next three years.	
I agree that I can be contacted after the end of this study about possible future research and follow-up with PD-CPA and related groups.	
I agree that my GP can be told that I am participating in this study.	

GP's name: _____ Surgery: _____

Address: _____

MRI and Cognition:

I agree to have an MRI scan and I understand what will happen in the scan. ☐

I have had an MRI safety check and I am confident that there is no reason why I can't have a scan, such as a recent operation. ☐

I agree that my test results can be held by the Wellcome Trust and shared with other research groups, and I understand that this data will be anonymous and not contain any personal information. ☐

Genetics:

You do not have to agree to provide blood or saliva samples to take part in the research. You do not have to agree that any samples you do give can be stored for future testing.
By giving a sample, you consent to be contacted by BioResource about the possibility of joining their panel, but you are under no obligation to join BioResource.

I agree to give a sample of **blood and saliva** (delete as appropriate) for medical research and for details about me and any samples I provide to be kept on a secure database. I agree that BioResource, the study collaborator on genetics, can store my samples and can contact me to invite me to join their panel. ☐

I agree that the samples and information I provide can be stored for use in future medical research, subject to ethical approval. ☐

I understand that I will not benefit financially if my samples are used in research leading to a new treatment or medical test being developed. ☐

In the unlikely event that an abnormality is picked up from tests carried out on my sample, I agree to be informed, and with my consent my GP can be told. ☐

Thank you for your help.

By completing and returning this form, you are giving us your consent that the personal information you provide will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.

Participant:
Signed: Date:

Researcher:

Signed: Date:

Appendix B5: Debrief form

Version 1.0 [Debriefing Sheet]

PD – CPA

Personality Disorders – a Computational Psychiatry Approach

Understanding the Social Brain in Healthy Volunteers and People with Psychological Difficulties.

Thank you for taking part in our study, we appreciate that you gave up your time to take part and hope that you found it interesting.

Summary of the Research Project

The aim of our study is to understand how mind and brain work in order to better understand patients with psychological difficulties. We hope that this will have an impact on the development of specific treatment interventions.

Most of our tasks are designed to look at how we think about ourselves and others (called "mentalisation"), how we regulate our emotions, value co-operation or experience close relationships and how problems can sometimes develop in these relationships.

Getting a better sense of the different strategies that people apply in these areas can help us understand more about when people experience mental health problems that can lead them to find certain social interactions and situations challenging. We hope to use these findings so that treatments can be tailored to help improve the domains where a patient's difficulties may lie.

We are also interested in how someone's experiences in childhood and his or her parenting at that time impact on the performances in the tasks and the functioning of the brain areas that underpin them. For instance, the long interview can tell us more about the quality of your bonding with parents.

Some of the topics discussed in the course of the study may have brought about thoughts or feelings which you had not previously considered or may have made you recall memories which could be perceived as distressing or lead you to feel tense or ruminate on thoughts. Therefore, we have provided some exercises at the back of this sheet which may help you to cope with any such feelings which you may experience.

What to do if you continue to feel concerned

If you continue to feel concerned after taking part in the study it may be useful to talk to a family member, a friend or your GP. Your Lead Clinician (care co-ordinator) or Probation Worker will also be able to support you, if you have one.

In addition to this support there is also free and confidential advice provided by the Mental Health charity Mind which can be found on their website: http://www.mind.org.uk/ or by calling their advice line 0300 123 3393.

If you feel at immediate risk do not hesitate to contact Dr Janet Feigenbaum (details overleaf).

wellcome trust UCLPartners UCL Anna Freud Centre

Contact Details

If you still have concerns or wish to contact the research team to discuss any of the information further or any concerns you have about the study, then please do so by getting in touch with the members of the research team listed below:

If you feel that we have not addressed your questions adequately or if you have any concerns about the conduct of the research team, then please contact my supervisor Dr. Janet Feigenbaum (Strategic and Clinical Lead for Personality Disorder Services, North East London NHS Foundation Trust and Senior Lecturer, Research Department of Clinical, Educational and Health Psychology, UCL) on 07957 919 961 or by email at janet.feigenbaum@nhs.net.

Janet Feigenbaum, PhD

Research Department of Clinical, Educational and Health Psychology

General Office, Room 436, 4th Floor
1-19 Torrington Place, London, WC1E 7HB

telephone: 07957 919 961

Tobias Nolte MD

Wellcome Trust Centre for Neuroimaging & Research Department of Clinical, Educational and Health Psychology

12 Queen Square

London

WC1N 3BG

Tobias.nolte@annafreud.org

Thank you very much for taking the time to read this information sheet.

Relaxation Exercises

Progressive Muscle Relaxation Technique
{Pause between instructions}

Begin by finding a comfortable position either sitting or lying down in a location where you will not be interrupted.
Allow your attention to focus only on your body. If you begin to notice your mind wandering, bring it back to the muscle you are working on.

Take a deep breath through your abdomen, hold for a few seconds, and exhale slowly. Again, as you breathe notice your stomach rising and your lungs filling with air.

As you exhale, imagine the tension in your body being released and flowing out of your body.
And again inhale.....and exhale. Feel your body already relaxing.

As you go through each step, remember to keep breathing .

Now let's begin. Tighten the muscles in your forehead by raising your eyebrows as high as you can. Hold for about five seconds. And abruptly release feeling that tension fall away.

Now smile widely, feeling your mouth and cheeks tense. Hold for about 5 seconds, and release, appreciating the softness in your face.

Next, tighten your eye muscles by squinting your eyelids tightly shut. Hold for about 5 seconds, and release.

Gently pull your head back as if to look at the ceiling. Hold for about 5 seconds, and release, feeling the tension melting away.

Now feel the weight of your relaxed head and neck sink.

Breath in...and out.

In...and out.

Let go of all the stress

In...and out.

Now, tightly, but without straining, clench your fists and hold this position until I say stop. Hold for about 5 seconds, and release.

Now, flex your biceps. Feel that buildup of tension. You may even visualize that muscle tightening.

Hold for about 5 seconds, and release, enjoying that feeling of limpness.

Breath in...and out.

Now tighten your triceps by extending your arms out and locking your elbows. Hold for about 5 seconds, and release.

Now lift your shoulders up as if they could touch your ears. Hold for about 5 seconds, and quickly release, feeling their heaviness.

Tense your upper back by pulling your shoulders back trying to make your shoulder blades touch.

Hold for about 5 seconds, and release.

Tighten your chest by taking a deep breath in, hold for about 5 seconds, and exhale, blowing out all the tension.

Now tighten the muscles in your stomach by sucking in. Hold for about 5 seconds, and release.

Gently arch your lower back. Hold for about 5 seconds, relax.

Feel the limpness in your upper body letting go of the tension and stress, hold for about 5 seconds, and relax.

Tighten your buttocks. Hold for about 5 seconds..., release, imagine your hips falling loose.

Tighten your thighs by pressing your knees together, as if you were holding a penny between them.

Hold for about 5 seconds...and release.

Mindfulness Exercise

Read the following instructions

Sit comfortably, with your eyes closed and your spine reasonably straight.

Bring your attention to your breathing.

Imagine that you have a balloon in your tummy. Every time you breathe in, the balloon inflates. Each time you breathe out, the balloon deflates. Notice the sensations in your abdomen as the balloon inflates and deflates. Your abdomen rising with the in-breath, and falling with the out-breath.

Thoughts will come into your mind, and that's okay, because that's just what the human mind does. Simply notice those thoughts, then bring your attention back to your breathing.

Likewise, you can notice sounds, physical feelings, and emotions, and again, just bring your attention back to your breathing.

You don't have to follow those thoughts or feelings, don't judge yourself for having them, or analyse them in any way. It's okay for the thoughts to be there. Just notice those thoughts, and let them drift on by, bringing your attention back to your breathing.

Whenever you notice that your attention has drifted off and is becoming caught up in thoughts or feelings, simply note that the attention has drifted, and then gently bring the attention back to your breathing.

It's okay and natural for thoughts to enter into your awareness, and for your attention to follow them. No matter how many times this happens, just keep bringing your attention back to your breathing.

Appendix B6: Measurement of symptoms modelled

	Symptom	Scale/ subscale	Cronbach's alpha*	No. items	Item scoring	Score range	Data type
ASOC							
1	Hostility	BSI subscale: hostility	0.872	5	Likert – 5 options (0-4)	0-20	Continuous
2	Aggression	LHA subscale: aggression	0.861	5	Likert – 6 options (0-5)	0-25	Continuous
3	Antisocial behaviour	LHA subscale: consequences/ antisocial	0.749	4	Likert – 6 options (0-5)	0-20	Continuous
4	Impulsive	APSD subscale: impulse/conduct problems	0.710	10	Likert – 3 options (0-2)	0-20	Continuous
5	Callous/ unemotional	APSD subscale: callous/ unemotional	0.517	6	Likert – 3 options (0-2)	0-12	Continuous
INT							
6	Somatising	BSI subscale: somatising	0.874	7	Likert – 5 options (0-4)	0-28	Continuous
7	Interpersonal sensitivity	BSI subscale: interpersonal sensitivity	0.905	4	Likert – 5 options (0-4)	0-16	Continuous
8	Depression	BSI subscale: depression	0.938	6	Likert – 5 options (0-4)	0-24	Continuous
9	Anxiety	BSI subscale: anxiety	0.920	6	Likert – 5 options (0-4)	0-24	Continuous
10	Specific phobia	BSI subscale: phobia	0.887	5	Likert – 5 options (0-4)	0-20	Continuous
11	Post-traumatic stress	PCL-S (complete scale)	0.958	17	Likert – 5 options (1-5)	17-85	Continuous
12	Obsessive compulsivity	BSI subscale: obsessive compulsive	0.905	6	Likert – 5 options (0-4)	0-24	Continuous
TD							
13	Psychoticism	BSI subscale: psychoticism	0.822	5	Likert – 5 options (0-4)	0-20	Continuous
14	Magic thinking	SPQ subscale: magic thinking	0.757	7	Dichotomous – N/Y (0/1)	0-7	Count**
15	Unusual perceptions	SPQ subscale: unusual perceptions	0.825	9	Dichotomous – N/Y (0/1)	0-9	Count**
16	Suspiciousness	SPQ subscale: suspicious	0.887	8	Dichotomous – N/Y (0/1)	0-8	Count**
17	Ideas of reference	SPQ subscale: ideas of reference	0.825	9	Dichotomous – N/Y (0/1)	0-9	Count**
18	Paranoid thoughts	BSI subscale: paranoid thinking	0.862	5	Likert – 5 options (0-4)	0-20	Continuous
19	Persecutory thoughts	GPTS (complete scale)	0.977	32	Likert – 5 options (1-5)	32-160	Continuous

Symptom		Scale/ subscale	Cronbach's alpha*	No. items	Item scoring	Score range	Data type
BOR							
20	Affective instability	PAI-BOR subscale: affective instability	0.817	6	Likert – 4 options (0-3)	0-18	Continuous
21	Emotional dysregulation	DERS (complete scale)	0.967	36	Likert – 5 options (1-5)	36-180	Continuous
22	Identity problems	PAI-BOR subscale: identity problems	0.790	6	Likert – 4 options (0-3)	0-18	Continuous
23	Negative relationships	PAI-BOR subscale: negative relationships	0.724	6	Likert – 4 options (0-3)	0-18	Continuous
24	Self-defeating behaviour	PAI-BOR subscale: self-injury	0.843	6	Likert – 4 options (0-3)	0-18	Continuous
25	Self-injury	LHA subscale: self-directed aggression	0.828	2	Likert – 6 options (0-5)	0-10	Continuous

Key: * calculated from recoded rather than raw scores, where relevant**; **** count data is left-censored and generally fits the Poisson distribution, however the distribution of these data were checked and judged to support them being treated as continuous in the analyses

Appendix B7: Table of missing data

Measure	Per cent missing at item-level
APSD	7.9%
BSI	6.4%
DERS	1.7%
GPTS	6.6%
LHA	1.6%
PAI	0.5%
PLC-S	2.0%
SPQ	1.0%
CTQ	7.7%
RFQ	6.7%
Total	**4.3%**

Appendix B8: Model specification details

Data

Full information maximum likelihood (FIML) is the default method for dealing with missing data in Mplus and therefore does not need to be specified. FIML 'skips' missing data and adjusts confidence in parameter estimates accordingly (Muthén & Muthén, 2006).

Estimator

Robust maximum likelihood (MLR) makes no assumptions about the distribution of variables in the population (Li, 2016), and uses tetrachoric correlation coefficients with a scaling factor to deal with non-normality (Muthén & Muthén, 2006)

Model

The variables were measured on different scales (Appendix B6) and so in order to aid model identification the default Mplus setting of identifying factor metrics by fixing the first loading was not used. Instead the first loading was freed and the factor variances were set to one (Geiser, 2013).

As all models with group factors were oblique, group factors were allowed to correlate. However, correlations between *p* and all group factors were set to zero in all bifactor models.

Output

All models were initially run with modification indices and standardised estimates as outputs. The best-fitting full bifactor and the comparator group factor models were then re-run, with factor scores saved as free format files (Muthén & Muthén, 2006), as shown in Appendix B9.

Appendix B9: Mplus syntax

Model 1: Unidimensional

```
TITLE: CFA unidimensional

DATA:
    FILE = "C:\Users\data.dat";

VARIABLE:
    NAMES = host agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si
    rfc rfu ctqpa ctqsa ctqea ctqpn ctqen ctq pd gender;
    USEVARIABLES = host agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si;
    MISSING = ALL(-999);

ANALYSIS:
    TYPE = GENERAL;
    ESTIMATOR = MLR;
    ITERATIONS = 1000;
    CONVERGENCE = 0.00005;

MODEL:
    p BY host* agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si;

    p@1;

OUTPUT: MOD STAND;
```

Model 2: Correlated group factors

```
TITLE: CFA correlated group factors

DATA:
    FILE = "C:\Users\data.dat";

VARIABLE:
    NAMES = host agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si
    rfc rfu ctqpa ctqsa ctqea ctqpn ctqen ctq pd gender;
    USEVARIABLES = host agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si;
    MISSING = ALL(-999);


ANALYSIS:
    TYPE = GENERAL;
    ESTIMATOR = MLR;
    ITERATIONS = 1000;
    CONVERGENCE = 0.00005;

MODEL:
    EXT BY host* agg as imp cu;
    INT BY oc* som is dep anx phob pts;
    TD BY psy* mag perc susp idea para pers;
    BOR BY ai* ed id nr sd si;

    EXT@1;
    INT@1;
    TD@1;
    BOR@1;

OUTPUT:
    STANDARDIZED MODINDICES (4);

SAVEDATA:
    FILE IS group.sav;
    SAVE IS fscores;
    FORMAT IS free;
```

Model 3: Full bifactor

```
TITLE: CFA full bifactor

DATA:
    FILE = "C:\Users\data.dat";

VARIABLE:
    NAMES = host agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si
    rfc rfu ctqpa ctqsa ctqea ctqpn ctqen ctq pd gender;
    USEVARIABLES = host agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si;
    MISSING = ALL(-999);


ANALYSIS:
    TYPE = GENERAL;
    ESTIMATOR = MLR;
    ITERATIONS = 1000;
    CONVERGENCE = 0.00005;

MODEL:
    p BY host* agg as imp cu som is dep anx phob pts oc
    psy mag perc susp idea para pers ai ed id nr sd si;

    EXT BY host* agg as imp cu;
    INT BY oc* som is dep anx phob pts;
    TD BY psy* mag perc susp idea para pers;
    BOR BY ai* ed id nr sd si;

    p@1;
    EXT@1;
    INT@1;
    TD@1;
    BOR@1;

    p WITH EXT @0;
    p WITH INT @0;
    p WITH TD @0;
    p WITH BOR @0;

OUTPUT: STANDARDIZED MODINDICES (4);

SAVEDATA: FILE IS bifactor.sav;
    SAVE IS fscores;
    FORMAT IS free;
```

Model 4: Modified bifactor with no TD group factor

```
TITLE: CFA modified bifactor - no TD

DATA:
    FILE = "C:\Users\data.dat";

VARIABLE:
    NAMES = host agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si
    rfc rfu ctqpa ctqsa ctqea ctqpn ctqen ctq pd gender;
    USEVARIABLES = host agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si;
    MISSING = ALL(-999);

ANALYSIS:
    TYPE = GENERAL;
    ESTIMATOR = MLR;
    ITERATIONS = 1000;
    CONVERGENCE = 0.00005;

MODEL:
    p BY host* agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si;

    EXT BY host* agg as imp cu;
    INT BY oc* som is dep anx phob pts;
    BOR BY ai* ed id nr sd si;

    p@1;
    EXT@1;
    INT@1;
    BOR@1;

    p WITH EXT @0;
    p WITH INT @0;
    p WITH BOR @0;


OUTPUT: MOD STAND;
```

Model 5: Modified bifactor with no BOR group factor

```
TITLE: CFA modified bifactor - no BOR

DATA:
    FILE = "C:\Users\data.dat";

VARIABLE:
    NAMES = host agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si
    rfc rfu ctqpa ctqsa ctqea ctqpn ctqen ctq pd gender;
    USEVARIABLES = host agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si;
    MISSING = ALL(-999);

ANALYSIS:
    TYPE = GENERAL;
    ESTIMATOR = MLR;
    ITERATIONS = 1000;
    CONVERGENCE = 0.00005;

MODEL:
    p BY host* agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si;

    EXT BY host* agg as imp cu;
    INT BY oc* som is dep anx phob pts;
    TD BY psy* mag perc susp idea para pers;

    p@1;
    EXT@1;
    INT@1;
    TD@1;

    p WITH EXT @0;
    p WITH INT @0;
    p WITH TD @0;

OUTPUT: MOD STAND;
```

Model 6: Modified bifactor with no INT group factor

```
TITLE: CFA modified bifactor - no INT

DATA:
    FILE = "C:\Users\data.dat";

VARIABLE:
    NAMES = host agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si
    rfc rfu ctqpa ctqsa ctqea ctqpn ctqen ctq pd gender;
    USEVARIABLES = host agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si;
    MISSING = ALL(-999);

ANALYSIS:
    TYPE = GENERAL;
    ESTIMATOR = MLR;
    ITERATIONS = 1000;
    CONVERGENCE = 0.00005;

MODEL:
    p BY host* agg as imp cu som is dep anx phob pts
    oc psy mag perc susp idea para pers ai ed id nr sd si;

    EXT BY host* agg as imp cu;
    TD BY psy* mag perc susp idea para pers;
    BOR BY ai* ed id nr sd si;

    p@1;
    EXT@1;
    TD@1;
    BOR@1;

    p WITH EXT @0;
    p WITH TD @0;
    p WITH BOR @0;

OUTPUT: MOD STAND;
```

References for Appendices

Geiser, C. (2013). *Data Analysis with Mplus* (1st edition).. New York: Guilford Press.

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936–949.

Muthén, L. K., & Muthén, B. O. (2006). Mplus Version 7 user's guide. *Los Angeles, CA: Muthén & Muthén*.